Documentación técnica

visión general y arquitectura de ECS

Resumen

En este documento se proporciona un resumen técnico y el diseño de la plataforma de almacenamiento de objetos a escala de nube definida por software Dell EMC™ ECS™.

Febrero 2021

H14071.18

Revisiones

Fecha	Descripción
Diciembre de 2015	Versión inicial
Mayo de 2016	Se actualizó para 2.2.1
Septiembre de 2016	Actualización para 3.0
Agosto de 2017	Actualización para 3.1
Marzo de 2018	Actualización para 3.2
Septiembre de 2018	Actualización para el hardware de Gen3
Febrero de 2019	Actualización para 3.3
Septiembre de 2019	Actualización para 3.4
Febrero de 2020	Cambios actualizados en ECSDOC-628
Mayo de 2020	Actualización para 3.5
Noviembre de 2020	Actualización para 3.6
Febrero de 2021	Actualización para 3.6.1

Agradecimientos

La producción de esta documentación estuvo a cargo de las siguientes personas:

Autor: Zhu, Jarvis

La información de esta publicación se proporciona “tal cual”. Dell Inc. no se hace responsable ni ofrece garantía de ningún tipo con respecto a la información de esta publicación y desconoce específicamente toda garantía implícita de comerciabilidad o capacidad para un propósito determinado. El uso, la copia y la distribución de cualquier software descrito en esta publicación requieren una licencia de software correspondiente.
Esta guía puede contener ciertas palabras que no son coherentes con las pautas de lenguaje actuales de Dell. Dell tiene planes para actualizar esta guía en versiones futuras posteriores para revisar estas palabras según corresponda.
Este documento puede contener lenguaje de contenido de terceros que no está bajo el control de Dell y no es coherente con las reglas actuales de Dell para el contenido de Dell. Cuando el tercero pertinente actualice el contenido de terceros, este documento se revisará según corresponda.
Copyright © 2015–2021 Dell Inc. o sus filiales. Todos los derechos reservados. Dell, EMC, Dell EMC y otras marcas comerciales son marcas comerciales de Dell Inc. o sus filiales. Las demás marcas comerciales pueden ser marcas comerciales de sus respectivos dueños. [22/10/2021] [Documentación técnica] [H14071.18]

Tabla de contenido

Resumen ejecutivo

Las organizaciones necesitan opciones para consumir servicios de nube pública con la confiabilidad y el control de una infraestructura de nube privada. Dell EMC ECS es una plataforma de almacenamiento de objetos definida por software, compatible con IPv6 y a escala de nube que ofrece servicios de almacenamiento S3, Atmos, CAS, Swift, NFSv3 y HDFS en una sola plataforma moderna.

Con ECS, los administradores pueden administrar de manera sencilla la infraestructura de almacenamiento distribuida globalmente bajo un único espacio de nombres global con acceso al contenido desde cualquier lugar. Los componentes principales de ECS están en capas para brindar flexibilidad y resiliencia. Cada capa se abstrae y se puede escalar de forma independiente con una alta disponibilidad.

Los desarrolladores están adoptando el acceso simple de la API RESTful para los servicios de almacenamiento. El uso de la semántica de HTTP, como GET y PUT, simplifica la lógica de la aplicación requerida en comparación con las operaciones de archivos tradicionales, pero conocidas, basadas en rutas. Además, el sistema de almacenamiento subyacente de ECS es altamente coherente, lo que significa que puede garantizar una respuesta autorizada. Las aplicaciones que se requieren para garantizar la entrega autorizada de datos son capaces de hacerlo sin lógica de código complejo con solo usar ECS.

1 Introducción

En este documento, se proporciona una visión general de la plataforma de almacenamiento de objetos de Dell EMC ECS. Proporciona información detallada sobre la arquitectura de diseño de ECS y los componentes principales, como los servicios de almacenamiento y los mecanismos de protección de datos.

1.1 Público

Este informe está destinado a cualquier persona interesada en comprender el valor y la arquitectura de ECS. Se propone brindar contexto con enlaces a información adicional.

1.2 Alcance

Este documento se centra principalmente en la arquitectura de ECS. No abarca los procedimientos de instalación, administración y actualización para el hardware o el software de ECS. Además, no cubre detalles específicos sobre el uso y la creación de aplicaciones con las API de ECS.

Las actualizaciones de este documento se realizan periódicamente y, por lo general, coinciden con las versiones principales o con las nuevas funciones.

2 Valor de ECS

ECS ofrece un valor significativo para las empresas y los proveedores de servicios que buscan una plataforma diseñada para apoyar el rápido crecimiento de los datos. Entre las principales ventajas y funciones de ECS que permiten a las empresas administrar y almacenar contenido distribuido globalmente a escala se incluyen las siguientes:

- **Escala de nube**: ECS es una plataforma de almacenamiento de objetos para cargas de trabajo tradicionales y de última generación. La arquitectura definida por software que está organizada en capas de ECS promueve una escalabilidad ilimitada. Funciones destacadas:
  - Infraestructura de objetos distribuida globalmente
  - Capacidad de exabyte + escala sin límites de capacidad de pool de almacenamiento, clúster o entorno federado
  - No existe ningún límite en el número de objetos en un sistema, un espacio de nombres o un depósito
  - Eficiencia en cargas de trabajo de archivos pequeños y grandes sin límites de tamaño de objetos
- **Implementación flexible**: ECS tiene una flexibilidad inigualable con funciones como las siguientes:
  - Implementación del dispositivo
  - Implementación solo de software compatible con hardware estándar del sector certificado o personalizado
  - Compatible con múltiples protocolos: objetos (S3, Swift, Atmos, CAS) y archivos (HDFS, NFSv3)
  - Cargas de trabajo múltiples: aplicaciones modernas y archivado a largo plazo
  - Almacenamiento secundario para Data Domain Cloud Tier e Isilon con CloudPools
  - Rutas de actualización no disruptivas a los modelos ECS de la generación actual
- **Nivel empresarial**: ECS proporciona a los clientes más control de sus recursos de datos con almacenamiento de clase empresarial en un sistema seguro y compatible con funciones como las siguientes:
  - Datos en reposo (D@RE) con rotación de claves y administración de claves externa.
  - Comunicación cifrada entre sitios
  - Deshabilita los puertos 9101/9206 de manera predeterminada para permitir que las organizaciones cumplan con las políticas de cumplimiento de normas
  - Generación de informes, retención de registros basada en políticas y eventos, y reforzamiento de la plataforma para el cumplimiento de normas SEC 17a-4 (f), incluida la administración de retención avanzada, como la retención para asuntos legales y de auditoría, y la gobernanza mínima y máxima
  - Cumplimiento con las pautas de reforzamiento de la Guía de implementación técnica de seguridad (STIG) de la Agencia de Sistemas de Información de Defensa (DISA).
  - Autenticación, autorización y controles de acceso con Active Directory y LDAP
  - Integración con la infraestructura de monitoreo y alertas (SNMP traps y registro del sistema)
  - Funcionalidades empresariales mejoradas (multiusuario, monitoreo de la capacidad y alertas)
- **Reducción del TCO**: ECS puede reducir considerablemente el costo total de la propiedad (TCO) en relación con el almacenamiento tradicional y el almacenamiento de nube pública. Incluso ofrece un TCO más bajo que la cinta para retención a largo plazo. Entre las características, se incluyen las siguientes:
  - Espacio de nombres global
  - Rendimiento de archivos pequeños y grandes
  - Migración transparente de Centera
  - Completamente compatible con REST de Atmos
  - Menor sobrecarga de administración
  - Espacio físico del centro de datos pequeño
  - Alta utilización de almacenamiento

El diseño de ECS está optimizado para los siguientes casos de uso principales:

- **Aplicaciones modernas**: ECS está diseñado para el desarrollo moderno, como las aplicaciones web, móviles y de nube de última generación. El desarrollo de aplicaciones se simplifica con almacenamiento altamente coherente. Junto con el acceso de lectura/escritura simultáneo de múltiples sitios y de múltiples usuarios, a medida que la capacidad de ECS cambia y crece, los desarrolladores nunca necesitan volver a codificar sus aplicaciones.
- **Almacenamiento secundario**: ECS se utiliza como almacenamiento secundario para liberar el almacenamiento primario de los datos a los que se accede con poca frecuencia y, al mismo tiempo, también se puede acceder a ellos satisfactoriamente. Algunos ejemplos son productos de organización en niveles basados en políticas, como Data Domain Cloud Tier e Isilon CloudPools. Geodrive, una aplicación basada en Windows, brinda a los sistemas Windows acceso directo a ECS para almacenar datos.
- **Archiving protegido geográficamente**: ECS puede servir como una nube segura y accesible en las instalaciones para fines de archiving y retención a largo plazo. Utilizar ECS como nivel de archivo puede reducir las capacidades de almacenamiento primario significativamente. Para permitir una mejor eficiencia de almacenamiento para los casos de uso de archivo inactivo, se encuentra disponible un esquema de codificación de eliminación (EC) de 10+2, además del valor predeterminado de 12+4.
- **Repositorio de contenido global**: los repositorios de contenido no estructurado que contienen datos como imágenes y videos suelen almacenarse en sistemas de almacenamiento de alto costo, lo que hace que sea imposible para los negocios administrar de manera rentable el crecimiento masivo de los datos. ECS permite la consolidación de múltiples sistemas de almacenamiento en un solo repositorio de contenido eficiente y accesible globalmente.
- **Almacenamiento para la Internet de las cosas**: la Internet de las cosas (IoT) ofrece una nueva oportunidad de ingresos para los negocios que pueden extraer valor de los datos del cliente. ECS ofrece una arquitectura de IoT eficiente para la recopilación de datos no estructurados a escala masiva. Sin límites en la cantidad de objetos, el tamaño de los objetos o los metadatos personalizados, ECS es la plataforma ideal para almacenar los datos de la IoT. ECS también puede optimizar algunos flujos de trabajo analíticos, ya que permite que los datos se analicen directamente en la plataforma de ECS sin necesidad de procesos de extracción, transformación y carga (ETL) que consumen mucho tiempo. Los clústeres de Hadoop pueden ejecutar consultas utilizando los datos almacenados en ECS mediante otra API de protocolo, como S3 o NFS.
- **Repositorio de pruebas de videovigilancia**: a diferencia de los datos de la IoT, los datos de videovigilancia tienen un conteo de almacenamiento de objetos mucho más pequeño, pero un espacio físico de capacidad mucho mayor por archivo. Aunque la autenticidad de los datos es importante, la retención de datos no es tan importante. ECS puede ser un área de descarga de bajo costo o una ubicación de almacenamiento secundario para estos datos. El software de administración de video puede aprovechar las funcionalidades enriquecidas de metadatos personalizados para etiquetar archivos con detalles importantes, como la ubicación de la cámara, el requisito de retención y el requisito de protección de datos. Además, los metadatos se pueden usar para establecer el archivo
en un estado de solo lectura a fin de garantizar una cadena de protección en el archivo.
- **Lagos de datos y análisis**: los datos y el análisis se han convertido en un diferenciador competitivo y en una fuente principal de generación de valor para las organizaciones. Sin embargo, la transformación de los datos en un valioso recurso corporativo es un tema complejo que puede implicar fácilmente el uso de docenas de tecnologías, herramientas y entornos. ECS proporciona un conjunto de servicios para ayudar a los clientes a recopilar, almacenar, controlar y analizar datos a cualquier escala.

3 Arquitectura

ECS está diseñado con unos pocos principios de diseño principales, por ejemplo, un espacio de nombres global con coherencia sólida, funcionalidad de escalamiento horizontal y multiusuario seguro, y un rendimiento superior para los objetos grandes y pequeños. ECS está diseñado como un sistema completamente distribuido que sigue el principio de las aplicaciones en la nube, donde cada función del sistema se crea como una capa independiente. Con este diseño, cada capa es escalable horizontalmente en todos los nodos del sistema. Los recursos se distribuyen entre todos los nodos para aumentar la disponibilidad y compartir la carga.

En esta sección, se detallará la arquitectura y el diseño de ECS en cuanto a software y hardware.

3.1 Descripción general

ECS se implementa en un conjunto de hardware estándar calificado del sector o como un dispositivo de almacenamiento listo para usar. Los componentes principales de ECS son los siguientes:

- **Portal y servicios de aprovisionamiento de ECS**: WebUI y CLI basadas en API para el autoservicio, la automatización, la generación de informes y la administración de nodos de ECS. Esta capa también maneja licencias, autenticación, multiusuario y servicios de aprovisionamiento, como la creación de un espacio de nombres.
- **Servicios de datos**: servicios, herramientas y API para soportar el acceso de objetos y archivos al sistema.
- **Motor de almacenamiento**: servicio principal responsable de almacenar y recuperar datos, administrar transacciones y proteger y replicar datos de manera local y entre sitios.
- **Fabric**: servicio de agrupación en clústeres para la administración del estado, la configuración y la actualización, y la generación de alertas.
- **Infraestructura**: utiliza SUSE Linux Enterprise Server 12 como sistema operativo base para el dispositivo listo para usar o los sistemas operativos Linux calificados para la configuración de hardware estándar del sector.
- **Hardware**: un dispositivo listo para usar o hardware estándar del sector calificado.

Figura 1 muestra una vista gráfica de estas capas, que se describen en detalle en las secciones siguientes.

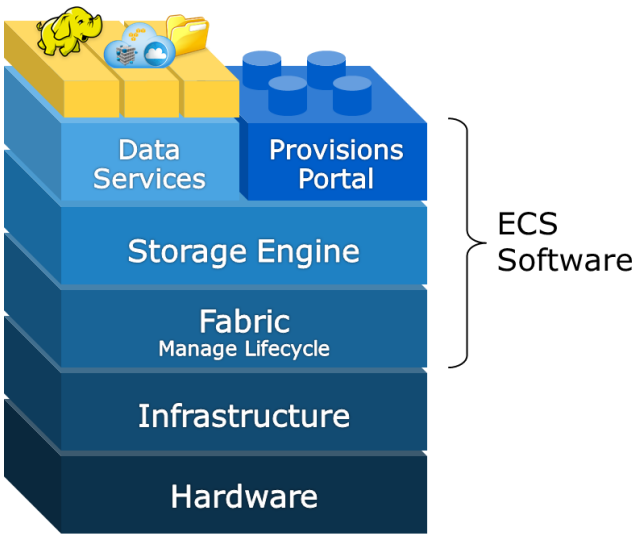

Figura 1 Capas de la arquitectura de ECS

3.2 Portal y servicios de aprovisionamiento de ECS

Los administradores de almacenamiento gestionan ECS mediante el portal y los servicios de aprovisionamiento de ECS. ECS proporciona una interfaz gráfica del usuario basada en la web (WebUI) para administrar, otorgar licencias y aprovisionar nodos de ECS. El portal tiene funcionalidades integrales de creación de informes que incluyen las siguientes:

- Utilización de la capacidad por sitio, pool de almacenamiento, nodo y disco
- Monitoreo de rendimiento de la latencia, el rendimiento y el progreso de replicación.
- Información de diagnóstico, como el estado de recuperación del nodo y el disco.

El tablero de ECS proporciona información general sobre el estado y el rendimiento en el nivel del sistema. Esta vista unificada mejora la visibilidad general del sistema. Las alertas notifican a los usuarios sobre eventos críticos, como límites de capacidad, límites de cuotas, fallas de discos o nodos, o fallas de software. ECS también proporciona una interfaz de línea de comandos para instalar, actualizar y monitorear ECS. El acceso a los nodos para el uso de la línea de comandos se realiza a través del protocolo SSH.
En Figura 2, que aparece a continuación, se muestra una captura de pantalla del tablero de ECS

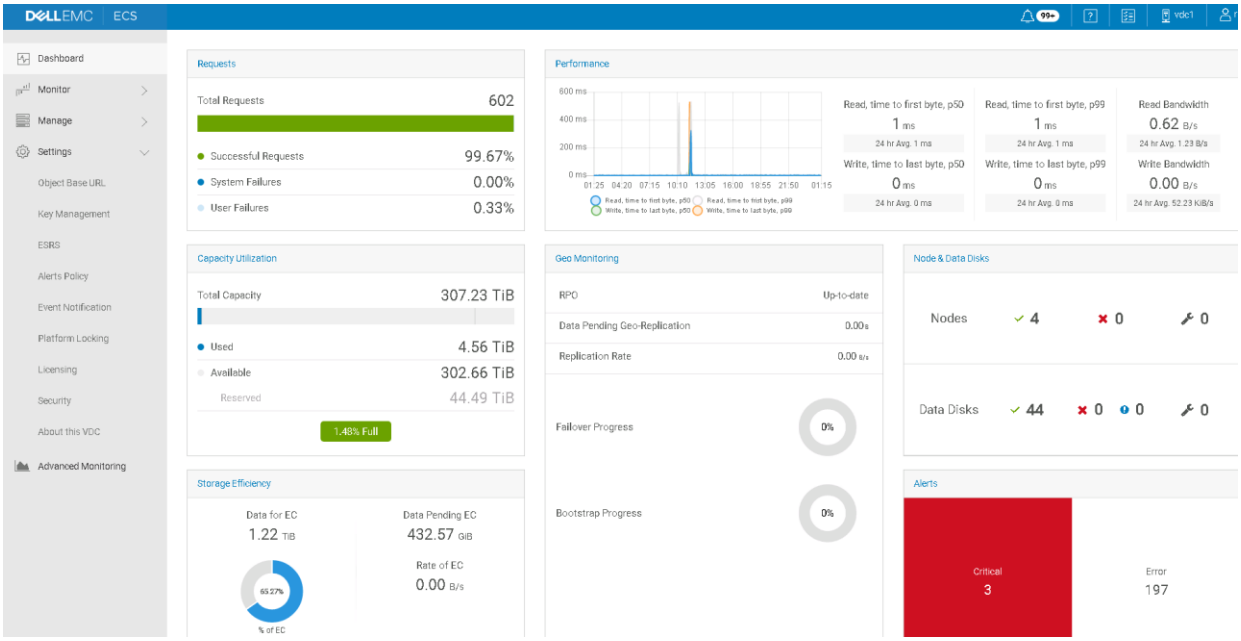

Figura 2 Tablero de la interfaz del usuario web de ECS

La creación de informes de rendimiento detallados está disponible en la UI, en la carpeta Monitoreo avanzado. Los informes se muestran en un tablero de Grafana. Hay filtros disponibles para desglosar a espacios de nombres, protocolos o nodos especificados. A continuación, se muestra un ejemplo de un informe de rendimiento del protocolo S3 en Figura 3.

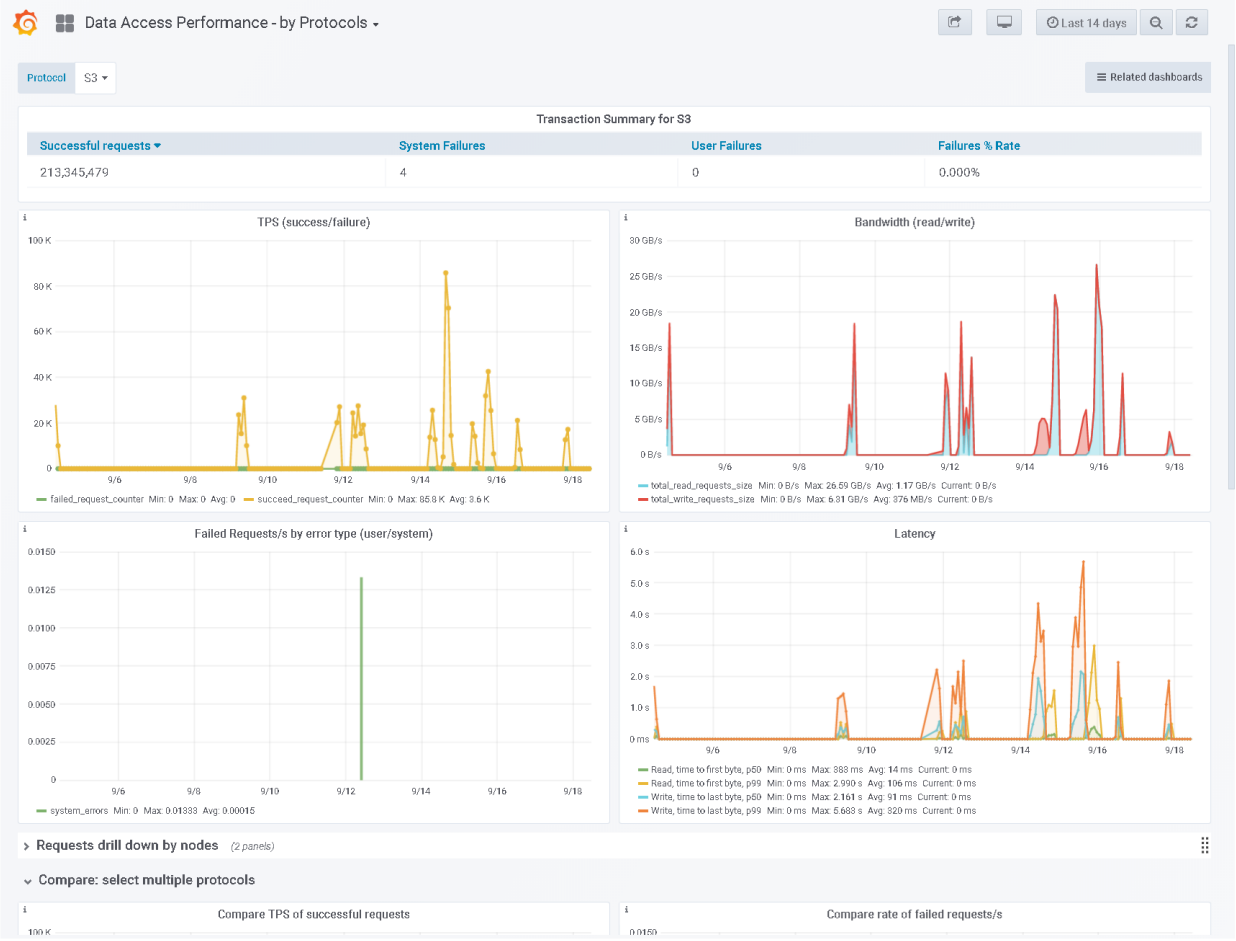

Figura 3 Visualización de monitoreo avanzado mediante Grafana

ECS también se puede administrar mediante las API RESTful. La API de administración permite a los usuarios administrar ECS desde sus propias herramientas, scripts y aplicaciones nuevas o existentes. Las herramientas de la interfaz del usuario Web y de la línea de comandos de ECS se crean mediante las API REST de administración de ECS.

ECS es compatible con los siguientes servidores de notificación de eventos, los cuales se pueden configurar mediante la interfaz del usuario Web, la API o la CLI:

- Servidores SNMP (Simple Network Management Protocol)
- Servidores de syslog

La *Guía del administrador de ECS* contiene más información y detalles sobre la configuración de los servicios de notificación.

3.3 Servicios de datos

Se usan métodos de objetos y archivos estándar para acceder a los servicios de almacenamiento de ECS. En el caso de S3, Atmos y Swift, se usan las API RESTful por HTTP para el acceso. Para el almacenamiento accedido por contenido (CAS), se utiliza un método de acceso/SDK de propiedad. ECS es compatible de forma nativa con todos los procedimientos de NFSv3, excepto LINK. S3a ahora puede acceder a los depósitos de ECS.

ECS proporciona acceso multiprotocolo en el que se puede acceder mediante un protocolo a los datos que se recopilan mediante otro protocolo. Esto significa que los datos se pueden recopilar a través de S3 y modificar a través de NFSv3 o Swift, o viceversa. Existen algunas excepciones a este acceso multiprotocolo, debido a la semántica del protocolo y las representaciones de diseño del protocolo. En Tabla 1 se destacan los métodos de acceso y los protocolos que interoperan.

Tabla 1 Interoperabilidad de protocolos y servicios de datos compatibles con ECS

Protocolos		**Compatible**	**Interoperabilidad**
Objetos	S3	Funcionalidades adicionales, como actualizaciones de rango de bytes y ACL enriquecidas	HDFS, NFS, Swift
	Atmos	Versión 2.0	NFS (solo objetos basados en rutas y no según el estilo de ID de objeto)
	Swift	Autenticación de las API de v2, Swift y Keystone v3	HDFS, NFS, S3
	CAS	SDK v3.1.544 o versiones posteriores	N/D
Archivo	HDFS	Compatibilidad con Hadoop 2.7	S3, NFS, Swift
	NFS	NFSv3	S3, Swift, HDFS, Atmos (solo objetos basados en rutas y no según el estilo de ID de objeto)

Los servicios de datos, que también se denominan servicios principales, son los responsables de realizar solicitudes de clientes, de extraer la información requerida y de transferirla al motor de almacenamiento para mayor procesamiento. Todos los servicios principales se combinan con un único proceso, *dataheadsvc,* que se ejecuta dentro de la capa de infraestructura. Este proceso se encapsula aún más dentro de un contenedor docker denominado *object-main.* que se ejecuta en todos los nodos de ECS. La sección *Infraestructura* de este documento aborda Docker de manera más detallada. Los requisitos de puertos de servicio de protocolo de ECS, como el puerto 9020 para la comunicación de S3, están disponibles en la *Guía de configuración de seguridad de ECS* más reciente.

3.3.1 Objetos

ECS es compatible con las API de CAS, S3, Atmos y Swift para el acceso a objetos. A excepción de CAS, los objetos o los datos se escriben, se recuperan, se actualizan y se eliminan a través de llamadas HTTP o HTTPS de GET, POST, PUT, DELETE y HEAD. Para CAS, se utiliza la comunicación estándar de TCP, al igual que métodos y llamadas de acceso específicos.

ECS proporciona una funcionalidad para la búsqueda de metadatos de objetos mediante un lenguaje de consulta enriquecido. Esta es una función eficaz de ECS que permite a los clientes de objetos de S3 buscar

objetos dentro de los depósitos mediante el sistema y los metadatos personalizados. Aunque la búsqueda se puede realizar utilizando cualquier metadato, ECS puede devolver consultas más rápido mediante la búsqueda de metadatos que se configuraron específicamente para indexarse en un depósito, especialmente en el caso de depósitos con miles de millones de objetos.

Se pueden indexar hasta treinta campos de metadatos definidos por usuario por depósito. Los metadatos se especifican en el momento de la creación del depósito. La función de búsqueda de metadatos se puede habilitar en los depósitos que tienen habilitado el cifrado del lado del servidor, sin embargo, los atributos de metadatos de usuario indexados que se utilizan como clave de búsqueda no se cifrarán.

Nota: existe un impacto en el rendimiento cuando se escriben datos en los depósitos configurados para indexar metadatos. El impacto en las operaciones aumenta a medida que aumenta el número de campos indexados. El impacto en el rendimiento se debe considerar cuidadosamente en el momento de elegir si se deben indexar los metadatos de un depósito y, de ser así, cuántos índices se deben mantener.

Para los objetos de CAS, la API de consulta de CAS proporciona una capacidad similar para buscar objetos en función de los metadatos que se mantienen para los objetos de CAS que no se deben habilitar explícitamente.

Para obtener más información sobre las API y las API de ECS para la búsqueda de metadatos, consulte la *Guía de acceso a datos de ECS* más reciente. Para los SDK de Atmos y S3, consulte el SDK de servicios de datos de Dell EMC del sitio GitHub o Dell EMC ECS. En el caso de CAS, consulte el sitio de la comunidad de Centera. Puede acceder a varios ejemplos, recursos y asistencia para los desarrolladores en la comunidad de ECS.

Las aplicaciones del cliente como el navegador S3y Cyberduck proporcionan una manera de probar o acceder a los datos almacenados en ECS rápidamente. La prueba de producto de ECS la proporciona Dell EMC, que permite el acceso a un sistema ECS orientado al público con fines de pruebas y desarrollo. Después del registro para la unidad de prueba de ECS, los terminales de REST se proporcionan con la información de identificación para cada uno de los protocolos de objetos. Cualquier persona puede usar la unidad de prueba de ECS para probar su aplicación de API de S3.

Nota: solo el número de metadatos que se pueden indexar por depósito se limita a treinta en ECS. No hay ninguna limitación en cuanto a la cantidad total de metadatos personalizados que se almacenan por objeto, solo la cantidad indexada para la búsqueda rápida.

3.3.2 HDFS

ECS puede almacenar datos de sistema de archivos Hadoop Como un sistema de archivos compatible con Hadoop, las organizaciones pueden crear repositorios de big data en ECS que la analítica de Hadoop puede consumir y procesar. El servicio de datos HDFS es compatible con Apache Hadoop 2.7, con soporte para ACL detallados y atributos de sistema de archivos extendidos.

ECS se validó y probó con Hortonworks (HDP 2.7). ECS también es compatible con YARN, MapReduce, Pig, Hive/Hiveserver2, HBase, Zookeeper, Flume, Spark y Sqoop.

3.3.2.1 Compatibilidad con Hadoop S3A

ECS es compatible con el cliente hadoop S3A para almacenar datos de Hadoop. S3A es un conector de código abierto para Hadoop, basado en el SDK oficial de Amazon Web Services (AWS). Se creó para abordar el escalamiento de almacenamiento y los problemas de costo que muchos administradores de

Hadoop estaban teniendo con HDFS. Hadoop S3A conecta los clústeres de Hadoop a cualquier almacén de objetos compatible con S3 que se encuentre en la nube pública, la nube híbrida o las instalaciones.

Nota: la compatibilidad de S3A está disponible en Hadoop 2.7 o versiones posteriores.

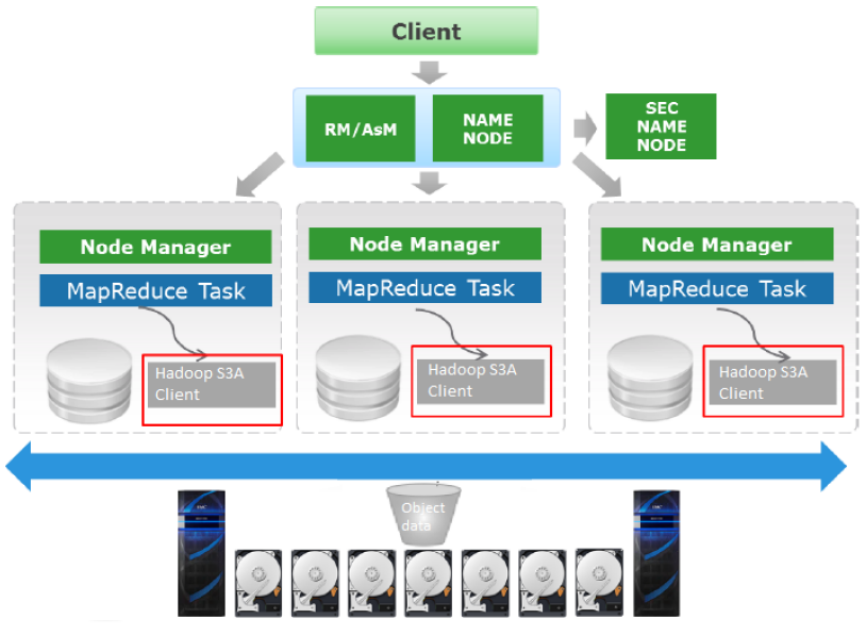

Figura 4 Arquitectura de Hadoop y ECS

Como se muestra en Figura 4, cuando el cliente configura el clúster de Hadoop en HDFS tradicional, su configuración de S3A apunta a los datos de objetos de ECS para realizar toda la actividad de HDFS. En cada nodo de Hadoop HDFS, cualquier componente tradicional de Hadoop usaría el cliente S3A de Hadoop para realizar la actividad de HDFS.

Análisis de la configuración de Hadoop mediante la consola de servicios de ECS

La consola de servicios (SC) de ECS puede leer e interpretar los parámetros de su configuración de Hadoop con respecto a las conexiones a ECS de S3A. Además, SC proporciona una función, *Get_Hadoop_Config* que lee la configuración del clúster de Hadoop y comprueba la configuración de S3A en busca de errores tipográficos, errores y valores. Póngase en contacto con el equipo de soporte de ECS para obtener ayuda con la instalación de ECS SC.

Implementación de Privacera con Hadoop S3A

Privacera es un proveedor tercero que implementó un agente del lado del cliente de Hadoop e integración con Ambari para la seguridad granular de S3 (AWS y ECS). Aunque Privacera es compatible con Cloudera Distribution de Hadoop (CDH), Cloudera (otro proveedor externo) no es compatible con Privacera en CDH.

Nota: los usuarios de CDH deben utilizar los servicios de seguridad de IAM de ECS. Si desea obtener acceso seguro a S3A sin utilizar la IAM de ECS, comuníquese con el equipo de soporte.

Consulte la *Guía de acceso a datos de ECS* más reciente para obtener más información sobre la compatibilidad con S3A

Seguridad de Hadoop S3A

IAM de ECS permite que el administrador de Hadoop configure políticas de acceso para controlar el acceso a los datos de S3A Hadoop. Una vez que se definen las políticas de acceso, hay dos opciones de acceso de usuario para que los administradores de Hadoop configuren:

- Usuarios/grupos IAM
  - Crear grupos de IAM que se conectan a políticas
  - Crear usuarios de IAM que sean miembros de un grupo de IAM
- Aserciones de SAML (usuarios federados)
  - Crear funciones de IAM que se conectan a políticas
  - Configurar CrossTrustRelationship entre el proveedor de identidades (AD FS) y ECS que asignan grupos de AD a funciones de IAM

El administrador de ECS y el administrador de Hadoop deben trabajar en conjunto para definir previamente las políticas adecuadas. Los ejemplos ficticias que aparecen a continuación describen tres tipos de usuarios de Hadoop para los que crearemos políticas. Estos son:

- **Administrador de Hadoop:** realice todas las operaciones, excepto crear depósito y eliminar depósito
- **Usuario avanzado de Hadoop:** realice todas las operaciones, excepto crear depósitos, eliminar depósitos y eliminar objetos
- Usuario de solo lectura de Hadoop: solo lista y objetos de lectura

Para obtener más información sobre IAM de ECS, consulte IAM de ECS en la página 38.

3.3.2.2 Compatibilidad con cliente ECS HDFS

ECS se integra con Ambari, lo que le permite implementar fácilmente el archivo jar de clientes HDFS de ECS y especifica HDFS de ECS como el sistema de archivos predeterminado en un clúster de Hadoop. El archivo jar se instala en cada nodo dentro de un clúster de Hadoop participante. ECS proporciona la funcionalidad de almacenamiento y del sistema de archivos equivalente a lo que los nodos de datos y nombre hacen en una implementación de Hadoop. ECS optimiza el flujo de trabajo de Hadoop eliminando la necesidad de migración de datos a un DAS de Hadoop local o la creación de un mínimo de tres copias. La Figura 5 que aparece a continuación muestra el archivo jar del cliente HDFS de ECS instalado en cada nodo de procesamiento Hadoop y el flujo de comunicación general.

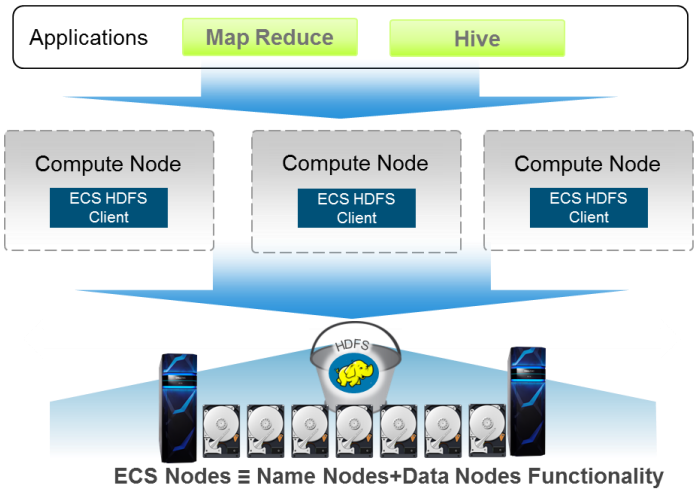

Figura 5 ECS que sirve de nodos de datos y nombre para un clúster de Hadoop

Entre otras mejoras agregadas en ECS para HDFS se incluyen las siguientes:

- **Autenticación de usuario de proxy**: suplantación de identidad de Hive, HBase y Oozie.
- **Seguridad**: aplicación de ACL del lado del servidor y adición de superusuario de Hadoop y grupo de superusuario, así como grupo predeterminado en depósitos.

3.3.3 NFS

ECS incluye soporte para archivos nativos con NFSv3. Entre las principales funciones del servicio de datos de archivos de NFSv3 se incluyen las siguientes:

- **Espacio de nombres global**: acceso a archivos desde cualquier nodo y sitio.
- **Bloqueo global**: en NFSv3, el bloqueo es **solo una recomendación**. ECS es compatible con implementaciones de clientes que cumplen con las normas que permiten bloqueos obligatorios y basados en rangos exclusivos y de uso compartido.
- **Acceso multiprotocolo**: acceso a datos mediante distintos métodos de protocolo.

Las exportaciones, los permisos y los mapeos de grupos de usuarios de NFS se crean mediante la interfaz del usuario Web o la API. Los clientes compatibles con NFSv3 montan las exportaciones mediante nombres de depósitos y espacio de nombres. Este es un comando de ejemplo para montar un depósito:

```
mount –t nfs –o vers=3 s3.dell.com:/namespace/bucket
```

Para lograr la transparencia del cliente durante una falla del nodo, se recomienda un balanceador de carga para este flujo de trabajo.

ECS integró estrechamente las otras implementaciones de servidores NFS, como *lockmgr*, *statd*, *nfsd*, and *mountd*, por lo tanto, estos servicios no dependen de la capa de infraestructura (sistema operativo del host) para llevar a cabo la administración. La compatibilidad de NFSv3 incluye las siguientes funciones:

- No hay límites de diseño en el número de archivos o directorios.
- El tamaño de escritura de archivos puede ser de hasta 16 TB.
- Capacidad de escalar en hasta ocho sitios con un único espacio de nombres o exportación global.
- Compatibilidad con la autenticación de Kerberos y de AUTH_SYS.

Los servicios de archivos NFS procesan las solicitudes de NFS que provienen de los clientes, sin embargo, los datos se almacenan como objetos dentro de ECS. Un identificador de archivo NFS se asigna a un ID de objeto. Dado que el archivo se asigna esencialmente a un objeto, NFS tiene funciones como el servicio de datos de objetos, que incluye lo siguiente:

- Administración de cuotas en el nivel del depósito.
- Cifrado en el nivel de objeto.
- Write-once-read-many (WORM) en el nivel del depósito.
  - WORM se implementa con el período de confirmación automática durante la creación del depósito nuevo.
  - WORM solo se aplica a los depósitos que no cumplen con las normas.

3.3.4 Conectores y gateways

Varios productos de software de otros fabricantes tienen la funcionalidad de acceder al almacenamiento de objetos de ECS. Proveedores de software independientes (ISV), como Panzura, Ctera y Syncplicity crean una capa de servicios que ofrece acceso de cliente al almacenamiento de objetos de ECS mediante

protocolos tradicionales, como SMB/CIFS, NFS e iSCSI. Las organizaciones también pueden acceder a los datos o cargarlos en el almacenamiento de ECS con los siguientes productos de Dell EMC:

- **Isilon CloudPools**: organización en niveles de datos basada en políticas en ECS desde Isilon.
- **Data Domain Cloud Tier**: organización en niveles nativa y automatizada de datos desduplicados en ECS desde Data Domain para la retención a largo plazo. Data Domain Cloud Tier ofrece una solución segura y rentable para cifrar los datos en la nube con espacio físico de almacenamiento reducido y ancho de banda de red.
- **GeoDrive**: servicio de almacenamiento basado en stubs de ECS servidores y escritorios de Microsoft® Windows®.

3.4 Motor de almacenamiento

El motor de almacenamiento es el núcleo de ECS. La capa del motor de almacenamiento contiene los componentes principales responsables del procesamiento de las solicitudes, así como el almacenamiento, la recuperación, la protección y la replicación de datos.

Esta sección describe los principios de diseño y cómo los datos se representan y manejan internamente.

3.4.1 Servicios de almacenamiento

El motor de almacenamiento de ECS incluye los siguientes servicios, como se muestra en la Figura 6.

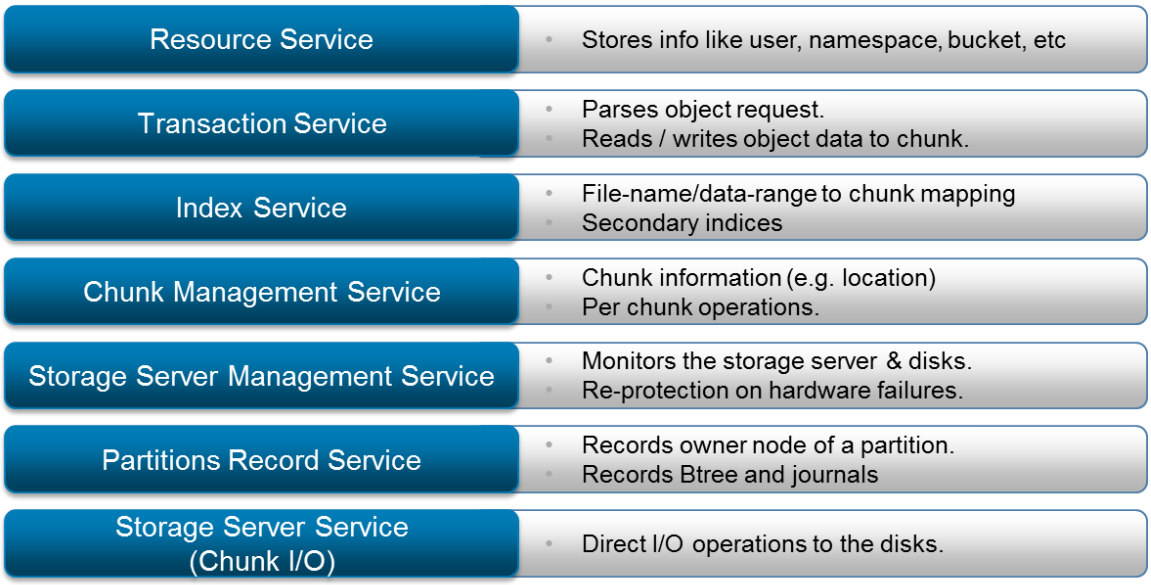

Figura 6 Servicios del motor de almacenamiento

Los servicios del motor de almacenamiento se encapsulan dentro de un contenedor de Docker que se ejecuta en cada nodo de ECS para proporcionar un servicio compartido y distribuido.

3.4.2 Datos

Los tipos principales de datos almacenados en ECS se pueden resumir de la siguiente manera:

- **Datos**: el contenido de la aplicación de datos o del nivel de usuario almacenado como una imagen. Los datos se utilizan como sinónimo de objetos, archivos o contenido. Las aplicaciones pueden almacenar una cantidad ilimitada de metadatos personalizados con cada objeto. El motor de almacenamiento escribe los datos y los metadatos personalizados proporcionados por la aplicación asociada en un repositorio lógico. Los metadatos personalizados son una función sólida de los sistemas de almacenamiento modernos que proporcionan más información o categorización de los datos que se almacenan. Los metadatos personalizados se formatean como pares de clave-valor y se proporcionan con solicitudes de escritura.
- **Metadatos del sistema**: información y atributos del sistema relacionados con los datos de usuario y los recursos del sistema. Los metadatos del sistema se pueden categorizar ampliamente de la siguiente manera:
  - **Identificadores y descriptores**: un conjunto de atributos que se utiliza internamente para identificar objetos y sus versiones. Los identificadores son ID numéricos o valores de hash que no se utilizan fuera del contexto del software de ECS. Los descriptores definen la información, como el tipo de codificación.
  - **Claves de cifrado en formato cifrado**: las claves de cifrado de datos se consideran metadatos del sistema. Se almacenan en un formato cifrado dentro de la estructura de la tabla del directorio principal.
  - **Marcas internas**: un conjunto de indicadores que se utilizan para rastrear si el cifrado o las actualizaciones de rango de bytes están habilitados, así como para coordinar el almacenamiento en caché y la eliminación.
  - **Información de ubicación**: atributo configurado con el índice y la ubicación de los datos, como las compensaciones de bytes.
  - **Registro de fecha y hora**: conjunto de atributos que rastrea la hora, como la creación o la actualización de un objeto.
  - **Información de configuración/grupos de usuarios**: control de acceso de objetos y espacio de nombres.

Los metadatos del sistema y los datos se escriben en *fragmentos* en ECS. Un fragmento de ECS es un contenedor lógico de espacio contiguo de 128 MB. Cada fragmento puede tener datos de diferentes objetos, como se muestra a continuación en la Figura 7. ECS utiliza la indexación para rastrear todas las partes de un objeto que se pueden distribuir entre diferentes fragmentos y nodos.

Los fragmentos están escritos en un patrón de solo anexo El comportamiento de solo anexo significa que la solicitud de una aplicación para modificar o actualizar un objeto existente no modificará ni eliminará los datos escritos anteriormente en un fragmento, sino que las nuevas modificaciones o actualizaciones se escribirán en un nuevo fragmento. Por lo tanto, no se requiere ningún bloqueo de I/O y no se requiere una invalidación de la caché. El diseño de solo anexo también simplifica el control de las versiones de los datos. Las versiones anteriores de los datos se mantienen en fragmentos anteriores. Si se habilita el control de versiones de S3 y se necesita una versión anterior de los datos, puede recuperarse o restaurarse a una versión anterior mediante la API REST de S3.

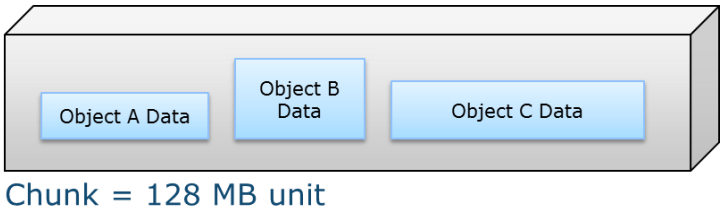

Figura 7 Fragmento de 128 MB que almacena datos de tres objetos

La sección *Protección e integridad de datos* a continuación explica cómo se protegen los datos en el nivel de fragmento.

3.4.3 Administración de datos

ECS utiliza un conjunto de tablas lógicas para almacenar información relacionada con los objetos. Los pares de clave-valor se almacenan finalmente en el disco en un árbol B+ para la indexación rápida de ubicaciones de datos. Mediante el almacenamiento de clave-valor en un árbol buscado y balanceado, como un árbol B+, se puede acceder rápidamente a la ubicación de los datos y los metadatos. ECS implementa un árbol de combinación con estructura de registros de dos niveles, donde hay dos estructuras similares a un árbol; un árbol más pequeño está en la memoria (tabla de memoria) y el árbol B+ principal reside en el disco. La búsqueda de pares de clave-valor se produce primero en la memoria y posteriormente en el árbol B+ principal en el disco, si es necesario. Las entradas en estas tablas lógicas primero se ingresan en los registros y estos se escriben en discos en fragmentos de triple espejeado. Los registros se utilizan para rastrear las transacciones que aún no se confirman en el árbol B+. Después de que cada transacción se ingresa en un registro, se actualiza la tabla en la memoria. Una vez que la tabla en la memoria se llena o después de un período determinado, su combinación se ordena o se vuelca en el árbol B+ en el disco. La cantidad de fragmentos de registro utilizados por el sistema es insignificante en comparación con los fragmentos del árbol B+. Figura 8 Ilustra este proceso.

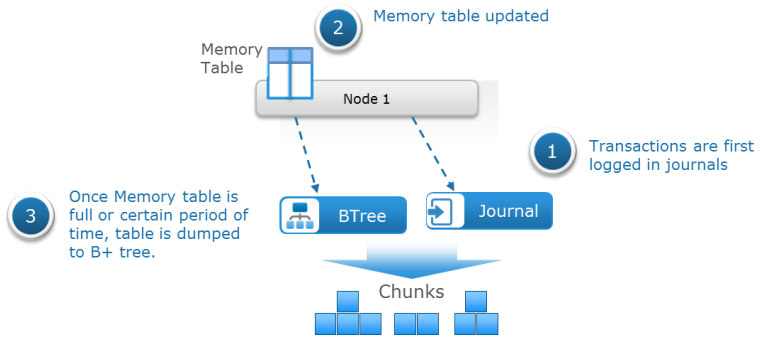

Figura 8 Tabla de memoria volcada en el árbol B+

La información almacenada en la tabla de objetos (OB) se muestra a continuación en la Tabla 2. La tabla de OB contiene los nombres de objetos y su ubicación de fragmento con una compensación y longitud determinadas dentro de ese fragmento. En esta tabla, el nombre del objeto es la clave del índice y el valor es la ubicación del fragmento. La capa de índice dentro del motor de almacenamiento es responsable de la asignación del nombre de objeto al fragmento.

Tabla 2 Entradas en la tabla de objetos

Nombre de objeto	Ubicación del fragmento
ImgA	• C1:offset:length
FileB	• C2:offset:length • C3:offset:length

La tabla de fragmentos (CT) registra la ubicación de cada fragmento como se muestra en la Tabla 3.

Tabla 3 Entradas en la tabla de fragmentos

ID del fragmento	Ubicación
C1	• Node1:Disk1:File1:Offset1:Length • Node2:Disk2:File1:Offset2:Length • Node3:Disk2:File6:Offset:Length

ECS se diseñó para ser un sistema distribuido, de modo que el almacenamiento y el acceso de los datos se distribuyan en todos los nodos. Las tablas que se utilizan para administrar los metadatos y los datos de objetos crecen con el tiempo, a medida que el almacenamiento se utiliza y crece. Las tablas se dividen en particiones y se asignan a diferentes nodos, y cada nodo se convierte en el propietario de las particiones que está alojando para cada una de las tablas. Por ejemplo, para obtener la ubicación de un fragmento, se consulta la tabla de registros de partición (PR) del nodo propietario que tiene la información de la ubicación del fragmento. En la Tabla 4 se muestra una tabla de PR básica.

Tabla 4 Entradas en la tabla de registros de partición

ID de partición	Propietario
P1	Nodo 1
P2	Nodo 2
P3	Nodo 3

Si un nodo deja de funcionar, otros nodos toman propiedad de sus particiones. Las particiones se recrean mediante la lectura de la raíz del árbol B+ y la reproducción de los registros almacenados en el disco. La Figura 9 muestra la conmutación por error de la propiedad de la partición.

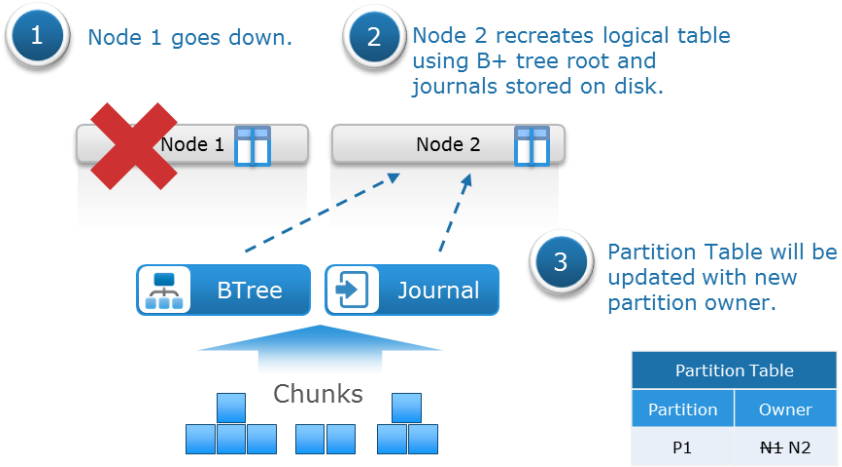

Figura 9 Conmutación por error de la propiedad de la partición

3.4.4 Flujo de datos

Los servicios de almacenamiento están disponibles desde cualquier nodo. Los datos están protegidos por segmentos de EC distribuidos en unidades, nodos y racks. ECS ejecuta una función de suma de comprobación y almacena el resultado con cada escritura. Si los primeros bytes de datos son comprimibles, ECS comprimirá los datos. Con las lecturas, los datos se descomprimen y se valida la suma de comprobación almacenada. Este es un ejemplo de un flujo de datos de una escritura en cinco pasos:

1. El cliente envía una solicitud de creación de objeto a un nodo.
2. El nodo que atiende la solicitud escribe los datos del nuevo objeto en un fragmento de repos. (abreviación de repositorio).
3. Cuando se escribe correctamente en el disco, se produce una transacción de PR para ingresar el nombre y la ubicación del fragmento.
4. El propietario de la partición registra la transacción en los registros.
5. Una vez que se registra la transacción en los registros, se envía una confirmación al cliente.

Como se muestra en Figura 10 a continuación, un ejemplo del flujo de datos para una lectura para la arquitectura de unidad de disco duro como Gen2 y EX300, EX500 y EX3000:

1. Se envía una solicitud de lectura de objeto del cliente al nodo 1.
2. El nodo 1 utiliza una función hash utilizando el nombre de objeto para determinar qué nodo es el propietario de la partición de la tabla lógica donde reside esta información de objeto. En este ejemplo, el nodo 2 es propietario y, por lo tanto, el nodo 2 llevará a cabo una búsqueda en las tablas lógicas para obtener la ubicación del fragmento. En algunos casos, la búsqueda puede ocurrir en dos nodos diferentes, por ejemplo, cuando la ubicación no se almacena en caché en las tablas lógicas del nodo 2.
3. En el paso anterior, la ubicación del fragmento se proporciona al nodo 1 que luego emitirá una solicitud de lectura de compensación de bytes al nodo que contiene los datos, el nodo 3 en este ejemplo, y enviará los datos al nodo 1.
4. El nodo 1 envía los datos al cliente que los solicitó.

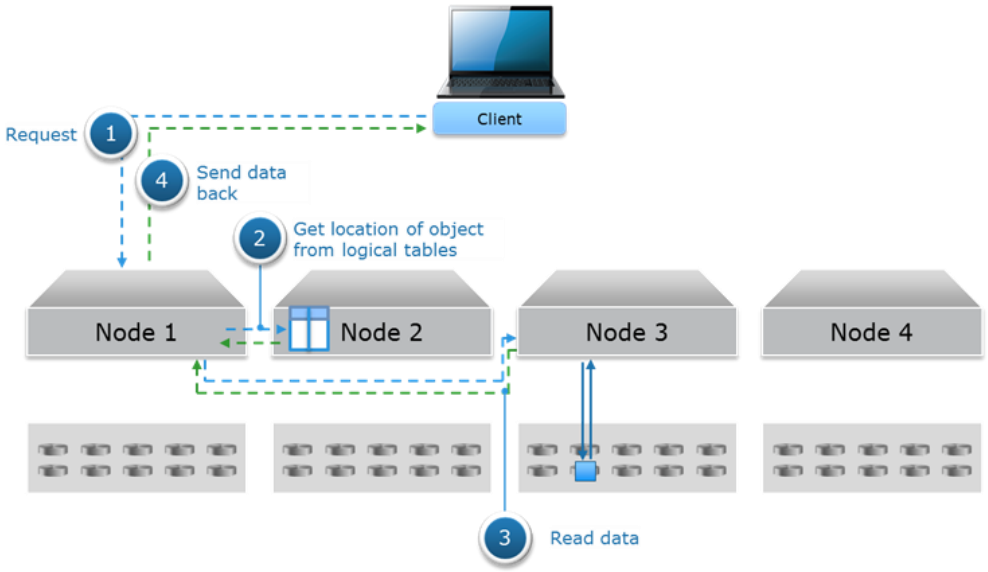

Figura 10 Flujo de datos de lectura para la arquitectura de la unidad de disco duro

Como se muestra en Figura 11 a continuación, un ejemplo del flujo de datos para una lectura para la arquitectura todo flash, como EXF900:

1. Se envía una solicitud de lectura de objeto del cliente al nodo 1.
2. El nodo 1 utiliza una función hash utilizando el nombre de objeto para determinar qué nodo es el propietario de la partición de la tabla lógica donde reside esta información de objeto. En este ejemplo, el nodo 2 es propietario y, por lo tanto, el nodo 2 llevará a cabo una búsqueda en las tablas lógicas para obtener la ubicación del fragmento. En algunos casos, la búsqueda puede ocurrir en dos nodos diferentes, por ejemplo, cuando la ubicación no se almacena en caché en las tablas lógicas del nodo 2.
3. En el paso anterior, la ubicación del fragmento se proporciona al nodo 1, que luego leerá los datos del nodo 3 directamente.
4. El nodo 1 envía los datos al cliente que los solicitó.

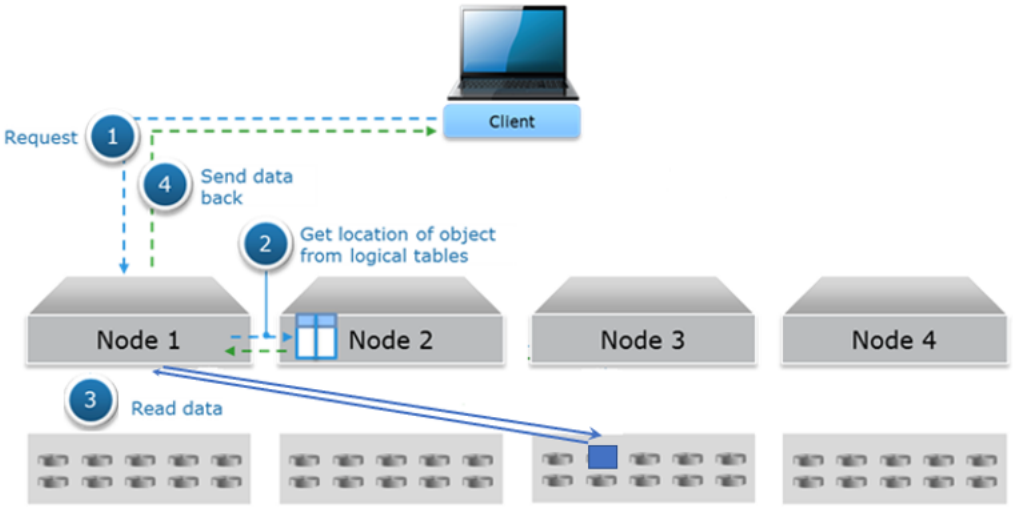

Figura 11 Flujo de datos de lectura para la arquitectura todo flash

Nota: en la arquitectura todo flash, como EXF900, cada nodo puede leer datos de otro nodo directamente, a excepción de la arquitectura de la unidad de disco duro en la que cada nodo solo puede leer el almacén de datos en sí mismo.

3.4.5 Optimizaciones de escritura para el tamaño de archivo

En el caso de escrituras en almacenamiento más pequeñas, ECS utiliza un método llamado *box-carting* para minimizar el impacto en el rendimiento. Box-carting agrega varias escrituras más pequeñas de 2 MB, o menos, en la memoria y las escribe en una sola operación de disco. Box-carting limita la cantidad requerida de recorridos de ida y vuelta al disco en el proceso de escritura individual.

En el caso de las escrituras de objetos más grandes, los nodos dentro de ECS pueden procesar las solicitudes de escritura del mismo objeto simultáneamente y aprovechar las escrituras simultáneas en varios ejes en el clúster de ECS. Por lo tanto, ECS puede recopilar y almacenar objetos pequeños y grandes de manera eficiente.

3.4.6 Recuperación de espacio

La escritura de fragmentos en una manera de solo anexo significa que los datos se agregan o se actualizan conservando, en primer lugar, los datos escritos originales y, en segundo lugar, mediante la creación de segmentos de fragmento nuevos netos que pueden o no incluirse en el contenedor de fragmentos del objeto original. El beneficio de la modificación de datos de solo anexo es un modelo de acceso de datos activo-activo que no se ve obstaculizado por problemas de bloqueo de archivos de sistemas de archivos tradicionales. Por este motivo, a medida que los objetos se actualizan o eliminan, ya no se hace referencia a los datos en fragmentos ni se necesitan. Los dos métodos de recolección de elementos no utilizados que utiliza ECS para recuperar espacio de fragmentos completos descartados, o fragmentos que contienen una combinación de partes de objetos eliminados y no eliminados a los cuales ya no se hace referencia, son los siguientes:

- **Recolección de elementos no utilizados normal**: cuando un fragmento completo está compuesto de elementos no utilizados, recupera espacio.
- **Recolección de elementos no utilizados parcial por combinación**: cuando los 2/3 de un fragmento son elementos no utilizados, recupera el fragmento combinando las partes válidas de este con otros fragmentos parcialmente llenos y creando un nuevo fragmento para recuperar espacio.

La recolección de elementos no utilizados también se aplica a la API de acceso de servicios de datos de ECS CAS para limpiar los blobs huérfanos. Los blobs huérfanos, que son blobs sin referencia identificados en los datos de CAS almacenados en ECS, serán elegibles para la recuperación de espacio a través de los métodos normales de recolección de elementos no utilizados.

3.4.7 Almacenamiento en caché de metadatos de SSD

Los metadatos de ECS se almacenan en árboles B. Cada árbol B puede tener entradas en la memoria, las transacciones de registro y en el disco. Para que el sistema tenga una imagen completa de un árbol B en particular, se consultan las tres ubicaciones, lo que a menudo incluye varias miradas al disco.

Para minimizar la latencia de las búsquedas de metadatos, se implementó un mecanismo de caché opcional basado en SSD en ECS 3.5. La caché contiene páginas de árbol B a las que se accedió recientemente. Esto significa que las operaciones de lectura en los árboles B más recientes siempre alcanzarán la caché basada en SSD y evitarán los viajes a discos giratorios.

Estos son algunos puntos destacados de la nueva función de almacenamiento en caché de metadatos de SSD:

- Latencia de lectura mejorada en todo el sistema y TPS (transacciones por segundo) para archivos pequeños
- Una unidad flash de 960 GB por nodo
- Los nuevos nodos netos de la fabricación incluyen la unidad SSD como una opción
- Los nodos de campo existentes (Gen3 y Gen2) se pueden actualizar a través de kits de actualización e instalación de autoservicio
- Las unidades SSD se pueden agregar mientras ECS está en línea
- Mejora para cargas de trabajo de análisis de archivos pequeños que requieren lecturas rápidas de grandes conjuntos de datos
- Todos los nodos de un VDC deben tener unidades SSD para habilitar esta función

El fabric de ECS detecta cuando se ha instalado un kit de SSD. Esto activa que el sistema se inicialice automáticamente y comience a utilizar la nueva unidad. Figura 12 muestra la caché de SSD habilitada.

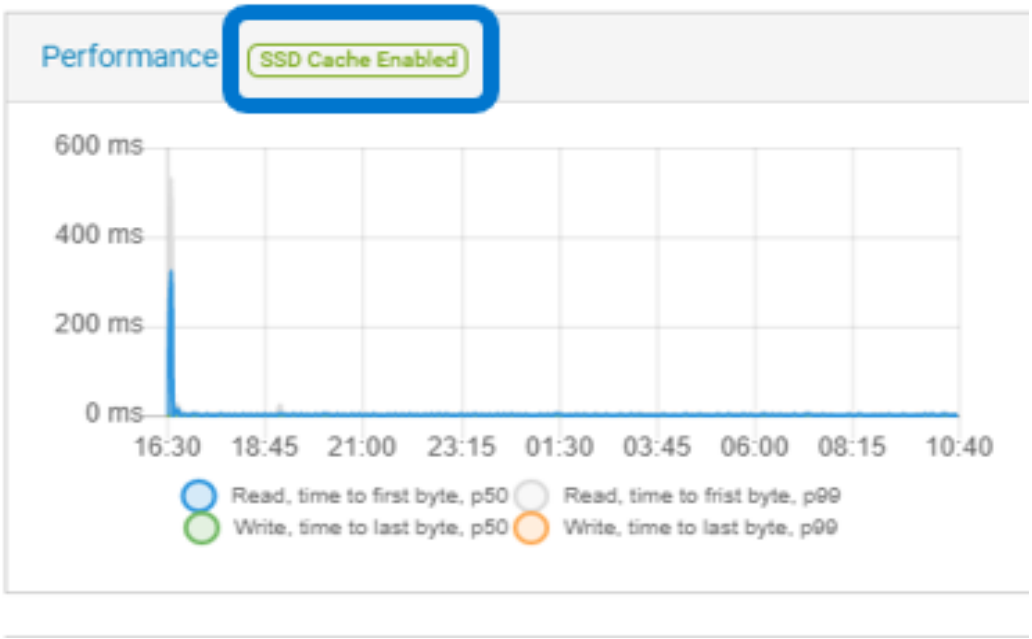

Figura 12 Caché de SSD habilitada

El almacenamiento en caché de metadatos de unidades SSD mejora las lecturas pequeñas y la lista de depósitos. Como se probó en nuestro laboratorio, el rendimiento de la lista mejora un 50 % con objetos de 10 MB. El rendimiento de lectura mejora un 35 % con objetos de 10 KB y un 70 % con objetos de 100 KB.

3.4.8 Cloud DVR

ECS es compatible con la función de grabación de video digital (DVR) de nube, la cual aborda un requisito legal de derechos de autor para las empresas de cables y televisión satelital. El requisito es que cada unidad de registro asignada a un objeto en ECS se debe copiar una cantidad de veces predeterminada. La cantidad predeterminada de copias se conoce como distribución. La cantidad predeterminada de copias (expansión) no es realmente un requisito de redundancia o aumento del rendimiento, sino que es más un requisito de derechos de autor legales para las empresas de cable y satélite. ECS es compatible con lo siguiente:

- Creación de una cantidad de copias de objetos creadas en ECS
- Permitir la lectura de una copia específica
- Permitir la eliminación de una copia específica
- Permitir la eliminación de todas las copias
- Permitir la copia de una copia específica
- Permitir la lista de copias
- Permitir la lista de depósitos de objetos de expansión

La función de DVR en la nube se puede habilitar a través de la consola de servicios. Por primera vez, debe habilitar la función DVR en la nube mediante la consola de servicio. Después de habilitar DVR en la nube, de manera predeterminada, para todos los nuevos nodos, DVR de en la nube está habilitado.

Ejecute el siguiente comando en la consola de servicios para habilitar la función DVR en la nube:

```
service-console run Enable_CloudDVR
```

La función de DVR en la nube admite API y puede consultar la *Guía de acceso a datos de ECS* para obtener más detalles

3.5 Fabric

La capa de Fabric proporciona agrupación en clústeres, estado del sistema, administración de software, administración de la configuración, funcionalidades de actualización y alertas. Es responsable de mantener los servicios en ejecución y administrar recursos, como los discos, los contenedores y la red. Rastrea los cambios en el ambiente, por ejemplo, la detección de una falla, y reacciona a ellos, y proporciona alertas relacionadas con el estado del sistema. La capa de Fabric tiene los siguientes componentes:

- **Agente de nodo**: administra los recursos de host (discos, red, contenedores, etc.) y los procesos del sistema.
- **Administrador del ciclo de vida**: administración del ciclo de vida de las aplicaciones que involucra el inicio de servicios, la recuperación, las notificaciones y la detección de fallas.
- **Administrador de persistencia**: coordina y sincroniza el entorno distribuido de ECS.
- **Registro**: almacenamiento de imágenes del Docker para el software de ECS.
- **Biblioteca de eventos**: contiene el conjunto de eventos que ocurren en el sistema.
- **Administrador de hardware**: proporciona el estado, la información de eventos y el aprovisionamiento de la capa de hardware a servicios de nivel superior. Estos servicios se integraron para ser compatibles con hardware genérico.

3.5.1 Agente de nodo

El agente de nodo es un agente ligero escrito en Java que se ejecuta de forma nativa en todos los nodos de ECS. Sus deberes principales incluyen la administración y el control de los recursos del host (contenedores de Docker, discos, el firewall y la red) y el monitoreo de los procesos del sistema. Algunos ejemplos de administración incluyen el formateo y el montaje de discos, la apertura de los puertos necesarios, la comprobación de la ejecución de todos los procesos y la determinación de interfaces de red pública y privada. Tiene un flujo de eventos que proporciona los eventos solicitados a un administrador del ciclo de vida para indicar eventos que ocurren en el sistema. Una CLI de Fabric es útil para diagnosticar problemas y observar el estado general del sistema.

3.5.2 Administración del ciclo de vida

El administrador del ciclo de vida se ejecuta en un subconjunto de tres o cinco nodos y administra el ciclo de vida de las aplicaciones que se ejecutan en los nodos. Cada administrador del ciclo de vida es responsable de rastrear varios nodos. Su objetivo principal es administrar todo el ciclo de vida de la aplicación de ECS desde el arranque hasta la implementación, incluida la detección de fallas, la recuperación, la notificación y la migración. Observa los flujos del agente de nodo e impulsa al agente para manejar la situación. Cuando un nodo está inactivo, responde a fallas o incoherencias en el estado del nodo mediante la restauración del sistema a un estado correcto conocido. Si una instancia del administrador del ciclo de vida está inactiva, otra toma su lugar.

3.5.3 Registro

El registro contiene las imágenes de Docker de ECS que se usan durante la instalación, la actualización y el reemplazo de nodos. Un contenedor de Docker llamado *fabric-registro* se ejecuta en un nodo dentro del rack de ECS y contiene el repositorio de imágenes de Docker de ECS y la información necesaria para las instalaciones

y las actualizaciones. A pesar de que el registro está disponible en un nodo a la vez, todas las imágenes de Docker se almacenan en caché localmente en cada nodo, por lo que cualquier otro puede servir al registro.

3.5.4 Biblioteca de eventos

La biblioteca de eventos se utiliza dentro de la capa de Fabric para exponer los flujos de eventos del agente del ciclo de vida y del nodo. Los eventos generados por el sistema se conservan en la memoria y el disco compartidos para proporcionar información histórica sobre el estado y la condición del sistema de ECS. Estos flujos de eventos solicitados se pueden usar para restaurar el sistema a un estado específico reproduciendo los eventos solicitados almacenados. Algunos ejemplos de eventos incluyen eventos de nodo como iniciado, detenido o degradado.

3.5.5 Administrador de hardware

El administrador de hardware se integra al agente de Fabric para admitir el hardware estándar del sector. Su propósito principal es proporcionar información sobre el estado y los eventos específicos del hardware, y el aprovisionamiento de la capa de hardware a servicios de nivel superior dentro de ECS.

3.6 Infraestructura

Los nodos de dispositivos de ECS actualmente ejecutan SUSE Linux Enterprise Server 12 para la infraestructura. Para el software de ECS implementado en un hardware estándar del sector personalizado, el sistema operativo también puede ser Linux RedHat Enterprise o CoreOS. Las implementaciones personalizadas se realizan a través de un proceso de validación y solicitud formal. Docker está instalado en la infraestructura para implementar las capas encapsuladas de ECS. El software de ECS está escrito en Java, por lo que la máquina virtual Java se instala como parte de la infraestructura.

3.6.1 Docker

ECS se ejecuta en el sistema operativo como una aplicación de Java y se encapsula dentro de varios contenedores de Docker. Los contenedores están aislados, pero comparten el hardware y los recursos del sistema operativo subyacente. Algunas partes del software de ECS se ejecutan en todos los nodos y algunas se ejecutan en solo uno o algunos nodos. Entre los componentes que se ejecutan dentro de un contenedor de Docker se incluyen los siguientes:

- **Objeto-principal**: contiene los recursos y procesos relacionados con los servicios de datos, el motor de almacenamiento, el portal y los servicios de aprovisionamiento. Se ejecuta en todos los nodos de ECS.
- **Fabric-ciclo de vida útil**: contiene los procesos, la información y los recursos necesarios para la administración del estado, la administración de la configuración y el monitoreo en el nivel del sistema. Siempre se ejecutará un número impar de instancias del fabric-ciclo de vida útil. Por ejemplo, habrá tres instancias en ejecución en un sistema de cuatro nodos y cinco instancias para un sistema de ocho nodos.
- **Fabric-Zookeeper**: servicio centralizado para la coordinación y sincronización de procesos distribuidos, información de configuración, grupos y servicios de asignación de nombres. Se conoce como administrador de persistencia y se ejecuta en un número impar de nodos, por ejemplo, cinco en un sistema de ocho nodos.
- **Fabric-registro**: registro de las imágenes de Docker de ECS. Solo se ejecuta una instancia por rack de ECS.

Hay otros procesos y herramientas que se ejecutan fuera de un contenedor de Docker, como el agente del nodo de Fabric y las herramientas de la capa de abstracción de hardware. La Figura 13 que aparece

a continuación proporciona un ejemplo de cómo se pueden ejecutar los contenedores de ECS en una implementación de ocho nodos.

Figura 13 Contenedores de Docker y agentes en el ejemplo de implementación de ocho nodos

La Figura 14 muestra el resultado de la línea de comandos del comando docker ps en un nodo, que muestra los cuatro contenedores que utiliza el Docker dentro de ECS. Se muestra una lista con todos los servicios relacionados con objetos disponibles en el sistema.

```
admin@hop-u300-11-pub-01:~> sudo docker ps
CONTAINER ID        IMAGE                                                COMMAND                  CREATED             STATUS
      PORTS               NAMES
7ba30ce42be2        ecs-monitoring/telegraf:3.5.0.0-825.b6b07cf9         "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-telegraf
e22513635cab        ecs-monitoring/grafana:3.5.0.0-825.b6b07cf9          "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-grafana
ee9db1ea40bc        emcvipr/object:3.5.0.0-120417.6a358e139f1            "/opt/vipr/boot/boot…"   5 weeks ago         Up 5 weeks
                          object-main
d11a7acd55e5        ecs-monitoring/throttler:3.5.0.0-825.b6b07cf9        "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-throttler
f94026797bb3        ecs-monitoring/fluxd:3.5.0.0-825.b6b07cf9            "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-fluxd
c7b8530a8bb9        caspian/fabric:3.5.0.0-4076.7d40a97                  "./boot.sh lifecycle"    5 weeks ago         Up 5 weeks
                          fabric-lifecycle
bffdf8836853        caspian/fabric-zookeeper:3.5.6.0-99.0354df7          "./boot.sh 1 1=169.2…"   5 weeks ago         Up 5 weeks
                          fabric-zookeeper
f4420f7f7d51        caspian/fabric-registry:2.3.1.0-68.10d1aca           "/opt/docker-registr…"   5 weeks ago         Up 5 weeks
                          fabric-registry
admin@hop-u300-11-pub-01:~> sudo dockobj
hop-u300-11-pub-01:/ # cd /opt/storageos/
hop-u300-11-pub-01:/opt/storageos # ls bin/*svc
bin/blobsvc     bin/coordinatorsvc  bin/eventsvc  bin/objcontrolsvc  bin/storagemanagementsvc
bin/cassvc      bin/dataheadsvc     bin/filesvc   bin/objheadsvc     bin/syssvc
bin/controlsvc  bin/ecsportalsvc    bin/hdfssvc   bin/resourcesvc    bin/transformsvc
```

Figura 14 Procesos, recursos, herramientas y archivos binarios en el contenedor objeto-principal

4 Modelos de hardware de dispositivo

Los puntos de entrada flexibles permiten a ECS escalar rápidamente a petabytes y exabytes de datos. Gracias a un impacto en el negocio mínimo, una solución de ECS puede escalar linealmente en capacidad y rendimiento mediante la adición de nodos y discos.

Los modelos de hardware de dispositivos de ECS se caracterizan por la generación de hardware. La serie de dispositivos de tercera generación, conocida como Gen3 o serie EX, incluye tres modelos de hardware. En esta sección, se proporciona una descripción general de alto nivel de la serie EX. Para obtener información detallada, consulte la *Guía de hardware de la serie EX de ECS*.

La información sobre el hardware de dispositivos de ECS de primera y segunda generación está disponible en la *Guía de hardware de la serie D y U de Dell EMC ECS*.

4.1 Serie EX

Los modelos de dispositivos de la serie EX se basan en servidores y switches estándares de Dell. Las ofertas de la serie son las siguientes:

- **EX300**: el EX300 tiene una capacidad cruda inicial de 60 TB. Es la plataforma de almacenamiento perfecta para las aplicaciones nativas de la nube y las iniciativas de transformación digital de los clientes. El EX300 es ideal para modernizar las implementaciones de Centera. Lo más importante es que EX300 puede escalar de manera rentable a capacidades más grandes. Proporciona 12 unidades por nodo y opciones de disco de 1 TB, 2 TB, 4 TB, 8 TB y 16 TB (todas iguales en el nodo)
- **EX500**: el EX500 es el dispositivo de edición más reciente que pretende proporcionar economía con densidad. Con opciones para 12 o 24 unidades, opciones de disco de 8 TB, 12 TB y 16 TB (todas iguales en el nodo). El rango del clúster va de 480 TB a 6,1 PB por rack. Esta serie proporciona una opción versátil para las medianas empresas que buscan soportar casos de uso de aplicaciones modernas o de archivo profundo.
- **EX3000**: el EX3000 cuenta con una capacidad máxima de 11,5 PB de almacenamiento crudo por rack, 30 a 90 unidades por nodo, discos de 12 TB o 16 TB y puede crecer en exabytes en varios sitios, lo que proporciona una solución de centro de datos profunda y escalable que es ideal para cargas de trabajo con un espacio físico de datos más grande. Estos nodos están disponibles en dos configuraciones diferentes conocidas como EX3000S y EX3000D. EX3000S es un solo nodo, y el EX3000D es un chasis de dos nodos. Estos nodos de alta densidad cuentan con intercambio de discos en caliente. Comienzan con un mínimo de treinta discos por nodo. Treinta unidades por nodo de ECS es el punto en el que disminuyen las mejoras de rendimiento tras la adición de más unidades. Con treinta o más unidades en cada nodo como mínimo, las expectativas de rendimiento son similares en cada nodo de EX3000, independientemente del conteo de unidades.
- **EXF900**: EXF900 es una solución de almacenamiento de objetos todo flash de nodos hiperconvergentes para implementaciones ECS de baja latencia y altos IOPS. Con opciones para 12 ó 24 unidades, opciones de unidad SSD NVMe de 3,84 TB (el controlador ssd NVMe de 7,68 TB será compatible cuando haya hardware disponible). Esta plataforma comienza con una configuración mínima cruda de 230 TB y escala hasta 1,4 PB de capacidad cruda por rack. Figura 15 muestra un nodo de EXF900.

Figura 15 Nodo de EXF900

Nota: La función de caché de lectura de SSD no se aplica a EXF900; Cloud DVR no es compatible con EXF900; Tech Refresh no es compatible con EXF900; EXF900 no puede coexistir con ningún otro hardware que no sea EXF900 en un VDC; EXF900 no puede coexistir con ningún otro hardware que no sea EXF900 en GEO (todos los sitios deben ser EXF900).

Las opciones de capacidad de la serie EX en el comienzo permiten a los clientes comenzar una implementación de ECS solo con la capacidad necesaria y crecer fácilmente a medida que cambian las necesidades en el futuro. Consulte la *Hoja de especificaciones de dispositivo de ECS* para obtener más detalles sobre los dispositivos de la serie EX, que también detalla los dispositivos de la serie U y D de la Gen2 anterior.

No se admiten actualizaciones posteriores a la implementación en los nodos de la serie EX. Estas incluyen:

- Cambio de CPU.
- Ajuste de la capacidad de la memoria.
- Actualización del tamaño del disco duro.

4.2 Redes del dispositivo

A partir de la versión de los dispositivos de la serie EX, se utiliza un par redundante de switches de administración de back-end dedicados. Mediante la migración al nuevo equipo de switches de dispositivos, ECS ahora puede adoptar un modo de conmutación de configuración de front-end y back-end.

Todos los dispositivos EX300, EX500 y EX3000 utilizan el S5148F de Dell EMC para el par de switches de front-end y el par de switches de back-end. El dispositivo EXF900 utiliza el S5248F de Dell EMC para el par de switches de front-end y para el par de switches de back-end y S5232F para el switch de back-end de agregación. Tenga en cuenta que los clientes tienen la opción de utilizar sus propios switches de front-end en lugar de los switches de Dell EMC.

4.2.1 S5148F: switches públicos de front-end

Se pueden obtener dos switches Ethernet S5148F opcionales de 1U, 25 GbE, de Dell EMC para la conexión de red, o bien el cliente puede proporcionar su propio par HA de 10 GbE o 25 GbE para la conectividad de front-end. A menudo, los switches públicos se conocen como hare y rabbit o solo el front-end.

Advertencia: es necesario contar con conexiones desde la red del cliente a ambos switches de front-end (Rabbit y Hare) a fin de mantener la arquitectura de alta disponibilidad del dispositivo de ECS. Si el cliente opta por no conectarse a su red de la forma de alta disponibilidad requerida, no existirá ninguna garantía de alta disponibilidad de datos para el uso de este producto.

Estos switches proporcionan 48 puertos de SFP28 de 25 GbE y 6 puertos de QSFP28 de 100 GbE. Otros detalles de estos dos tipos de puertos son los siguientes:

- SFP28 es una versión mejorada de SFP+
  - SFP+ soporta hasta 16 GB/s, SFP28 admite hasta 28 Gb/s
  - El mismo factor de forma
  - Compatible con versiones anteriores de módulos SFP+
- QSFP28 es una versión mejorada de QSFP+
  - QSFP+ soporta hasta 4 canales de 16 GB/s, QSFP28 soporta hasta 4 canales de 28Gb/s.
    - QSFP+ cuenta con canales agregados para obtener Ethernet de 40 GB/s
    - QSFP28 cuenta con canales agregados para obtener Ethernet de 100 GB/s
  - El mismo factor de forma
  - Compatible con versiones anteriores de módulos QSFP+
  - Se puede dividir en 4 canales individuales de SFP28

Nota: con los switches públicos S5148F de 25 GbE de Dell EMC, se proporcionan dos cables LAG de 100 GbE. Las organizaciones que proporcionan sus propios switches públicos deben proporcionar los cables de conexión de LAG, SFP o externos necesarios.

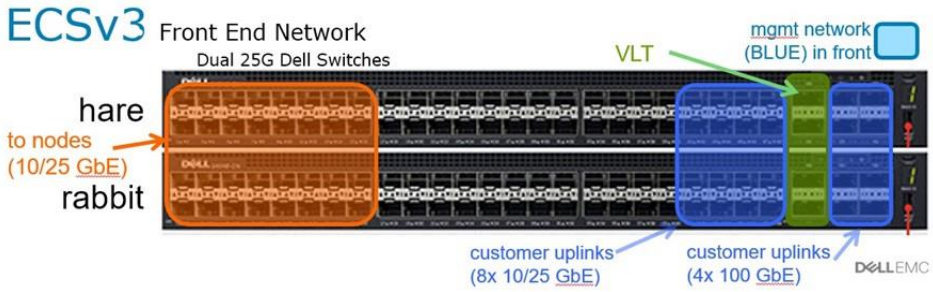

Figura 16 Designación y uso del puerto de switch de red de front-end

La Figura 16 anterior proporciona una representación visual de la manera en que los puertos están diseñados para permitir el tráfico de nodos de ECS, así como los puertos de enlace ascendente del cliente. Esto es estándar en todas las implementaciones.

4.2.2 S5148F: switches privados de back-end

Ambos switches Ethernet S5148F de 25 GbE y 1U de Dell EMC con 48 puertos SFP de 25 GbE y 6 puertos de enlace ascendente de 100 GbE necesarios se incluyen en cada rack de ECS. A menudo se denominan switches *fox* y *hound* o back-end y son responsables de la red de administración. En futuras versiones de ECS, los switches de back-end también proporcionarán la separación de la red para el tráfico de replicación. El propósito principal de la red privada es la consola y la administración remota, el arranque de PXE para el administrador de instalación, y habilitar el aprovisionamiento y la administración de rack y de todo el clúster. La Figura 17 muestra una vista frontal de dos switches de administración Dell de 25 GbE.

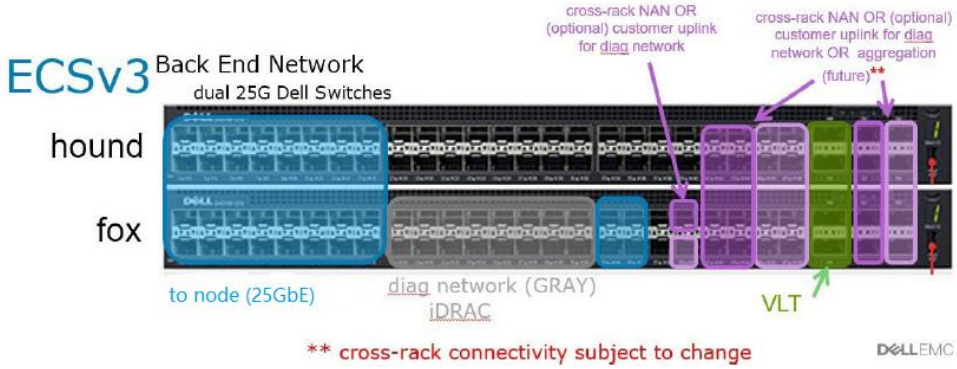

Figura 17 Designación y uso del puerto de switch de red de back-end

El diagrama anterior proporciona una representación visual de la manera en que los puertos están destinados a usarse para habilitar los puertos de diagnóstico y el tráfico de administración de ECS. Estas asignaciones de puertos son estándares en todas las implementaciones. Los posibles puertos de uso futuros se indican en púrpura, sin embargo, este uso está sujeto a cambios en el futuro.

4.2.3 S5248F: switches públicos de front-end

Dell EMC ofrece un par de alta disponibilidad opcional de switches S5248F de front-end de 25 GbE para la conexión de red del cliente al rack. Tiene dos cables de troncalización de enlace virtual (VLT) de 200 GbE (QSFP28-DD) por par de alta disponibilidad. Estos switches se denominan Hare y Rabbit. La Figura 18 muestra una representación visual de la manera en que los puertos están diseñados para permitir el tráfico de nodos de ECS, así como los puertos de enlace ascendente del cliente.

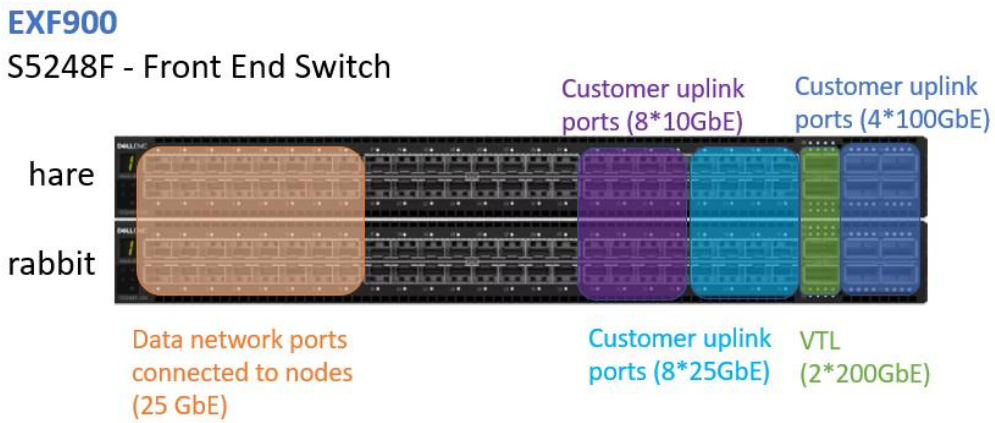

Figura 18 Designación y uso del puerto de switch de red de front-end

4.2.4 S5248F: switches privados de back-end

Dell EMC proporciona dos switches de back-end S5248F de 25 GbE con dos cables VLT de 200 GbE (QSFP28-DD). Estos switches se denominan switch Hound y switch Fox. Todos los cables iDRAC de los nodos y todas las conexiones del cable de administración del switch de front-end se dirigen al switch Fox. Figura 19 proporciona una representación visual de la manera en que los puertos están destinados a usarse para habilitar los puertos de diagnóstico y el tráfico de administración de ECS. Estas asignaciones de puertos son estándares en todas las implementaciones.

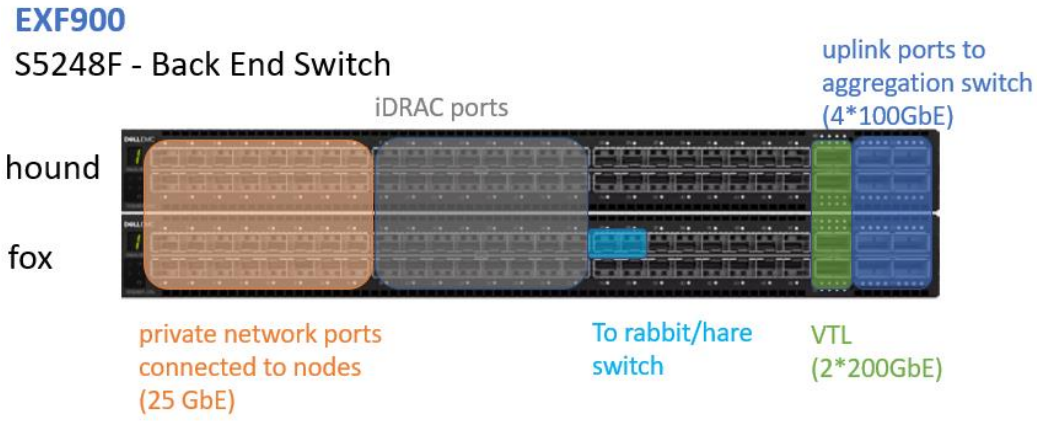

Figura 19 Designación y uso del puerto de switch de red de back-end

4.2.5 S5232: switch de agregación

Dell EMC proporciona dos switches de agregación de back-end S5232F de 100 GbE (AGG1 y AGG2) con cuatro cables VLT de 100 GbE. Estos switches se denominan switch Falcon y Eagle. En el siguiente Figura 20, todos los puertos etiquetados indican las designaciones de los puertos. Esta configuración permite conectar hasta 7 racks de nodos de EXF900.

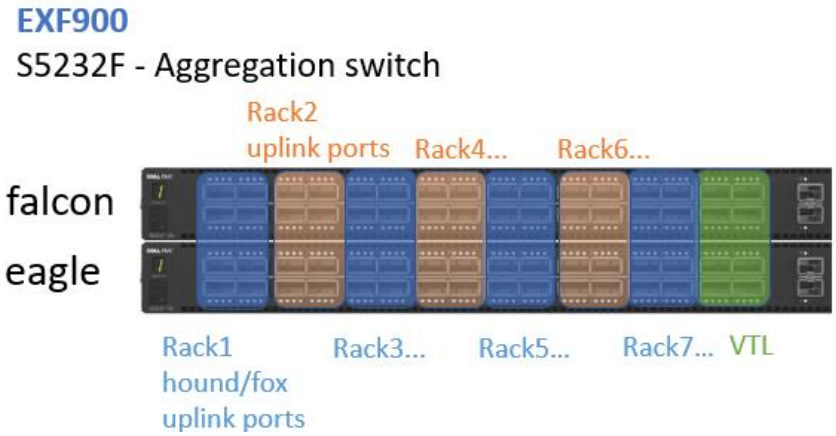

Figura 20 Designación y uso del puerto del switch de agregación

Para obtener más información sobre las redes y el cableado, consulte la *Guía de hardware de la serie EX de ECS.*

5 Separación de la red

ECS es compatible con la separación de diferentes tipos de tráfico de red para el aislamiento de la seguridad y el rendimiento. Los tipos de tráfico que se pueden separar incluyen los siguientes:

- Management
- Replicación
- Datos

Hay un modo de operación llamado *modo de separación de la red*. En este modo, cada nodo puede configurarse en el nivel del sistema operativo con un máximo de tres direcciones IP, o redes lógicas, para cada uno de los distintos tipos de tráfico. Esta función se diseñó para proporcionar la flexibilidad de crear tres redes lógicas separadas para la administración, la replicación y los datos, o la combinación de estas para crear dos redes lógicas, para la administración de instancias y el tráfico de replicación en una red lógica y tráfico de datos en otra red lógica. Se puede configurar una segunda red de datos lógica para el tráfico solo de CAS, lo cual permite la separación del tráfico de CAS de otros tipos de tráfico de datos, como S3.

La implementación de la separación de red de ECS requiere que cada tráfico de red lógica esté asociado a los servicios y los puertos. Por ejemplo, los servicios del portal de ECS se comunican a través de los puertos 80 o 443, por lo que estos puertos y servicios se vincularán a la red lógica de administración. Se puede configurar una segunda red de datos, sin embargo, es solo para el tráfico de CAS. La Tabla 5 que aparece a continuación destaca los servicios fijos para un tipo de red lógica. Para obtener una lista completa de los servicios asociados con los puertos, consulte la *Guía de configuración de seguridad de ECS* más reciente.

Tabla 5 Servicios para asignación de red lógica

Servicios	Red lógica	Identificador
Interfaz del usuario Web y API, SSH, DNS, NTP, AD y SMTP	Management	public.mgmt
Datos del cliente	Datos	public.data
	Datos de solo CAS	public.data2
Datos de replicación	Replicación	public.repl
Dell EMC Secure Remote Services (SRS)	El gateway de SRS está conectado según la red	public.data o public.mgmt

Nota: ECS 3.6 permite el acceso a datos de S3 tanto en la red de datos (predeterminada) como en la de data2 (aunque S3 no está habilitado de manera predeterminada en data2). Para habilitar el acceso a datos de S3 en la red data2, se requiere public.data y debe comunicarse con el soporte remoto de ECS.

La separación de la red se alcanza de manera lógica mediante distintas direcciones IP, virtualmente mediante la utilización de distintas VLAN o físicamente mediante cables diferentes. El comando *setrackinfo* se usa para configurar las direcciones IP y las VLAN. La configuración de VLAN de nivel de switch o del lado del cliente es responsabilidad del cliente. Para la separación de red física, los clientes deben enviar una solicitud de calificación de producto (RPQ) poniéndose en contacto con Dell EMC Global Business Services. Para obtener más información sobre la separación de redes, consulte la documentación técnica *Redes y prácticas recomendadas de ECS*, que proporciona una vista general de la separación de redes.

6 Seguridad

La seguridad de ECS se implementa en los niveles de administración, transporte y datos. La autenticación de usuarios y administradores se logra a través de Active Directory, los métodos de LDAP, Keystone o directamente en el portal de ECS. La seguridad de nivel de datos se realiza a través de HTTPS para los datos en movimiento o mediante cifrado del lado del servidor para los datos en reposo.

6.1 Autenticación

ECS es compatible con los métodos de autenticación de Active Directory, LDAP, Keystone e IAM para brindar acceso a la administración y la configuración de ECS. Sin embargo, existen limitaciones, como se muestra en la Tabla 6. Para obtener más información sobre seguridad, consulte la *Guía de configuración de seguridad de ECS* más reciente.

Tabla 6 Métodos de autenticación soportados

Método de autenticación	**Compatible**
Active Directory	• Soporte para grupos de AD como usuarios de administración • Soporte para grupos de AD para los métodos de autoaprovisionamiento de usuarios de objetos mediante el uso de claves de autoservicio a través de la API • Soporte para varios dominios
LDAP	• Los usuarios de administración pueden autenticarse individualmente mediante LDAP • Los grupos de LDAP NO son compatibles con los usuarios de administración • LDAP es compatible con los usuarios de objetos (claves de autoservicio a través de la API) • Soporte para varios dominios
Keystone	• Las políticas de RBAC aún no son compatibles. • No es compatible con tokens no incluidos en el alcance • No es compatible con varios servidores Keystone por sistema ECS
IAM	• Ofrece federación de identidades e inicio de sesión único (SSO) mediante estándares SAML 2.0 • Disponible solo a través del protocolo de S3

6.2 Autenticación de servicios de datos

El acceso a objetos mediante las API RESTful se protege a través de HTTPS (TLS v1.2). Las solicitudes entrantes se autentican mediante métodos definidos, como código de autenticación de mensajes basado en hash (HBAC), Kerberos o autenticación de token. La Tabla 7 que aparece a continuación presenta los diferentes métodos que se usan para cada protocolo.

Tabla 7 Autenticación de servicios de datos

Protocolos		**Métodos de autenticación**
Objetos	S3	V2 (HMAC-SHA1), V4 (HMAC-SHA256)
	Swift	Token: Keystone v2 y v3 (tokens PKI, UUID dentro del alcance), SWAuth v1
	Atmos	HMAC-SHA1
	CAS	Archivo PEA de clave secreta
Archivo	HDFS	Kerberos
	NFS	Kerberos, AUTH_SYS

6.3 Cifrado de datos en reposo (D@RE)

Los requisitos de cumplimiento de normas suelen exigir el uso del cifrado para proteger los datos escritos en los discos. En ECS, el cifrado se puede habilitar en los niveles de espacio de nombres y de depósito. Las funciones clave de D@RE de ECS incluyen las siguientes:

- Cifrado con baja intervención nativa en reposo: configuración simple y fácil habilitada
- Se utilizan las CLAVES (AES-256 CTR)
- Cifrado de clave pública de RSA con 2048 bits de longitud
- Soporte en el nivel de clúster de administración de claves externa (EKM):
  - SafeNet de Gemalto
  - Administrador del ciclo de vida de claves de seguridad de IBM
- Rotación de claves
- La semántica de cifrado S3 soporta el uso de encabezados HTTP como *x-amz-server-side-encryption*
- Cumplimiento de FIPS 140-2 con los estándares de seguridad criptográfica del gobierno de los EE. UU.

Nota: el modo FIPS 140-2 aplica el uso de algoritmos solo aprobados dentro de D@RE; el cumplimiento de normas de FIPS 140-2 es solo para el módulo D@RE, no para todo el producto ECS.

ECS utiliza una jerarquía de claves para cifrar y descifrar datos. El administrador de claves nativo almacena una clave privada común para todos los nodos con el fin de descifrar la clave principal. Con la configuración de EKM, la EKM proporciona la clave principal. Las claves proporcionadas por la EKM residen en la memoria únicamente en ECS. Nunca se guardan en el almacenamiento persistente dentro de ECS.

En un entorno replicado geográficamente, cuando un nuevo sistema ECS se une a una federación existente, la clave principal se extrae usando la clave pública-privada del sistema existente y se cifra utilizando el par de claves públicas-privadas nuevo que se generó en el nuevo sistema que se unió a la federación. A partir de este punto, la clave principal es global y conocida por ambos sistemas dentro de la federación. Cuando se usa la EKM, todos los sistemas federados recuperan la clave principal desde el sistema de administración de claves.

6.3.1 Rotación de claves

ECS soporta el cambio de claves de cifrado. Esto se puede hacer periódicamente para limitar la cantidad de datos protegidos por un conjunto específico de claves de cifrado de claves (KEK) o en respuesta a una posible pérdida o riesgo. Se utiliza un registro de KEK de rotación en combinación con otras claves de elemento primario para crear claves de encapsulación virtual con el fin de proteger las claves de cifrado de datos (DEK) y las KEK de espacio de nombres.

Las claves de rotación se generan o suministran de forma nativa y la EKM las mantiene. ECS utiliza la clave de rotación actual para crear claves de encapsulamiento virtual a fin de proteger cualquier DEK o KEK, independientemente de si la administración de claves se realiza de forma nativa o externa.

Durante las escrituras, ECS encapsula las DEK generadas aleatoriamente mediante una clave de encapsulación virtual que se crea utilizando el depósito y la clave de rotación activa.

Como parte de la rotación de claves, ECS vuelve a encapsular todos los registros de la KEK del espacio de nombres con una nueva KEK principal virtual creada a partir de la nueva clave de rotación, el contexto de la seña secreta asociada y la clave principal activa. Esto se realiza para proteger el acceso a los datos protegidos por las claves de rotación anteriores.

El uso de EKM afecta la ruta de lectura/escritura de los objetos cifrados. La rotación de claves permite la protección de datos adicional mediante el uso de claves de encapsulación virtual para las DEK y las KEK de espacio de nombres. Las claves de encapsulación virtual no persisten y se derivan de dos jerarquías independientes de claves persistentes. Con el uso de EKM, la clave de rotación no se almacena en ECS y aumenta aún más la seguridad de los datos. Principalmente agregamos nuevos registros de KEK y actualizamos los ID activos, pero nunca eliminamos nada.

Los puntos adicionales que se deben considerar con respecto a la rotación de claves en ECS son los siguientes:

- El proceso de rotación de claves solo cambia la clave de rotación actual. Las claves principales, de espacio de nombres y de depósito existentes no cambian durante el proceso de rotación de claves.
- La rotación de claves en el nivel de espacio de nombres o depósito no está soportada; sin embargo, el alcance de la rotación es en el nivel del clúster, por lo que todos los objetos nuevos cifrados del sistema se verán afectados.
- Los datos existentes no se vuelven a cifrar debido a las claves en rotación.
- ECS no soporta la rotación de claves durante un corte de suministro eléctrico.
  - TSO durante rotación: la tarea de rotación de claves se suspende hasta que el sistema sale de TSO.
  - PRO está en curso. ECS no debe estar en PSO para habilitar la rotación de claves. Si se produce un PSO durante la rotación, la rotación fallará de inmediato.
- No se requiere cifrado de depósitos para realizar el cifrado de objetos a través de S3.
- Los metadatos de objetos de cliente indexados que se utilizan como clave de búsqueda no se cifran.

Consulte la *Guía de configuración de seguridad de ECS* más reciente para obtener más información sobre D@RE, EKM y la rotación de claves.

6.4 IAM de ECS

La administración de acceso e identidades (IAM) de ECS permite tener control y acceso seguro a los recursos de S3 de ECS. Con esta funcionalidad, se garantiza que cada solicitud de acceso a un recurso de ECS se identifique, autentique y autorice. IAM de ECS permite al administrador agregar usuarios, funciones y grupos. El administrador también puede restringir el acceso mediante la adición de políticas a las entidades de IAM de ECS.

Nota: IAM de ECS se utiliza solo con S3. No está habilitado para los depósitos habilitados del sistema de archivos o CAS.

IAM de ECS está compuesto por los siguientes componentes

- **Administración de cuentas**: le permite administrar las identidades de IAM dentro de cada espacio de nombres, como usuarios, grupos y funciones.
- **Administración de acceso**: el acceso se administra mediante la creación de políticas y su conexión a recursos o identidades de IAM.
- **Federación de identidad**: SAML (Lenguaje de marcado de aserción de seguridad) establece y autentica la identidad. Una vez establecida la identidad, utilizará el servicio de token seguro para obtener las credenciales temporales que se utilizarán para acceder al recurso
- **Servicio de token seguro**: le permite solicitar credenciales temporales para el acceso entre cuentas a los recursos y también para los usuarios autenticados mediante la autenticación SAML de un proveedor de identidades empresariales o un servicio de directorio

Mediante IAM, puede controlar quién está autenticado y autorizado para usar los recursos de ECS mediante la creación y la administración

- **Usuarios**: un usuario de IAM constituye una persona o una aplicación en el espacio de nombres que puede interactuar con los recursos de ECS.
- **Grupos**: un grupo de IAM es una recopilación de usuarios de IAM. Utilice grupos para especificar permisos para una recopilación de usuarios de IAM.
- **Funciones**: función de IAM es una identidad que cualquier persona que requiera la función podría asumir. Una función es similar a un usuario, una identidad con políticas de permisos que determinan lo que la identidad puede hacer y lo que no.
- **Políticas**: una política de IAM es un documento en formato JSON que define los permisos para una función. Asigne y adjunte políticas a usuarios de IAM, grupos de IAM y funciones de IAM.
- **Proveedor SAML**: SAML es un estándar abierto para intercambiar datos de autenticación y autorización entre un proveedor de identidades y un proveedor de servicios. El proveedor de SAML en ECS se utiliza para establecer la confianza entre un proveedor de identidad compatible con SAML (IdP) y ECS

Cada sistema ECS se asigna con una cuenta de IAM de ECS. Esta cuenta es compatible con varios espacios de nombres y tiene entidades de IAM relacionadas que se definen en su espacio de nombres.

- Compatibilidad con espacios de nombres individuales en la administración de cuentas mediante las entidades de IAM de ECS, como usuarios, funciones y grupos.
- Las políticas, los permisos, la Lista de control de acceso (ACL) que están asociadas con las entidades de IAM de ECS y los recursos de S3 de ECS admiten la administración del acceso a las funciones de IAM de ECS.
- IAM de ECS es compatible con el acceso entre cuentas mediante el Lenguaje de marcado de aserción de seguridad (SAML) y las funciones.
- IAM de ECS es compatible con la clave de acceso de Amazon Web Services (AWS) para acceder a IAM y S3 en ECS.

Consulte la *Guía de seguridad de ECS* más reciente para obtener más información sobre IAM de ECS

6.5 Object tagging

El etiquetado de objetos permite categorizar los objetos mediante la asignación de etiquetas a los objetos individuales. Un único objeto puede tener varias etiquetas asociadas, lo que permite la categorización multidimensional.

Una etiqueta podría describir algún tipo de información confidencial, como un registro de estado, o puede etiquetar un objeto en un producto determinado que se puede categorizar como confidencial. El etiquetado es un subrecurso de un objeto que tiene un ciclo de vida útil integrado con operaciones de objetos. Puede agregar etiquetas a los objetos nuevos cuando los cargue, o bien a los objetos existentes. Es aceptable utilizar etiquetas para etiquetar objetos que contengan datos confidenciales, como información de identificación personal (PII) o información de estado protegida (PHI). Las etiquetas no deben contener información confidencial, ya que se pueden ver sin tener el permiso de lectura real de un objeto.

6.5.1 Información adicional sobre el etiquetado de objetos

En esta sección se proporciona información sobre el etiquetado de objetos en IAM, el etiquetado de objetos con políticas de depósito, el manejo del etiquetado de objetos durante TSO/PSO y el etiquetado de objetos durante la administración del ciclo de vida de los objetos. Estas son consideraciones adicionales:

- Etiquetado de objetos en IAM
  - La función clave del etiquetado de objetos como sistema de categorización se produce cuando se integra con políticas de IAM. Esto permite que el administrador configure permisos de usuario específicos. Por ejemplo, el administrador puede agregar una política que permita a todos los usuarios acceder a los objetos con una etiqueta especificada o puede configurar y otorgar permisos a los usuarios, quienes pueden administrar las etiquetas en objetos específicos. El otro aspecto clave con el etiquetado de objetos es la manera y la ubicación en que persisten las etiquetas. Esto es importante, debido a que tiene un impacto directo en diversos aspectos del sistema.
- Etiquetado de objetos con políticas de depósito
  - El etiquetado de objetos permite categorizar los objetos y además se integra con diversas políticas. La política de administración del ciclo de vida permite configurar en un nivel de depósito. Las versiones anteriores de ECS son compatibles con el vencimiento, las cargas de anulación incompletas y la eliminación del marcador de eliminación de etiquetado de objetos vencido. Es posible que el filtro incluya varias condiciones, incluida una condición basada en etiquetas. Cada etiqueta de la condición de filtro debe coincidir con la clave y el valor.
- Etiquetado de objetos durante TSO/PSO
  - El etiquetado de objetos es otro conjunto de entrada en los metadatos del sistema, no se requiere ningún manejo especial durante TSO/PSO. Existe un límite establecido para la cantidad de etiquetas que se pueden asociar a cada objeto, la extensión de los metadatos del sistema junto con el etiquetado de objetos se deben encontrar dentro de los límites de la memoria.
- Etiquetado de objetos durante la administración del ciclo de vida de objetos
  - El etiquetado de objetos es parte de los metadatos del sistema y se maneja simultáneamente con el manejo de metadatos del sistema durante la administración del ciclo de vida. La lógica de vencimiento y el escáner de eliminación del ciclo de vida útil requieren comprender políticas basadas en etiquetas. Las etiquetas de objetos permiten la administración del ciclo de vida de los objetos en detalle, en la cual puede especificar un filtro basado en etiquetas, además de un prefijo de nombre de clave, en una regla de ciclo de vida.

Consulte la *Guía de seguridad de ECS* más reciente para obtener más información sobre el etiquetado de objetos de ECS.

7 Integridad y protección de datos

Para la integridad de datos, ECS utiliza sumas de comprobación. Las sumas de comprobación se crean durante las operaciones de escritura y se almacenan con los datos. Durante las lecturas, las sumas de comprobación se calculan y se comparan con la versión almacenada. Un análisis de tareas en segundo plano verifica proactivamente la información de la suma de comprobación.

Para la protección de datos, ECS utiliza el espejeado triple con los fragmentos de registro y esquemas de EC independientes para los fragmentos de *repos.* (datos de repositorio de usuario) y *btree* (árbol B+).

La codificación de eliminación permite una protección de datos mejorada ante una falla de disco, de nodo y de rack en términos de eficiencia de almacenamiento en comparación con los esquemas de protección convencionales. El motor de almacenamiento de ECS implementa la corrección de errores de Reed Solomon utilizando los dos esquemas siguientes:

- 12+4 (valor predeterminado): el fragmento se divide en 12 segmentos de datos. Se crean 4 segmentos de codificación (paridad).
- 10+2 (archivo inactivo): el fragmento se divide en 10 segmentos de datos. Se crean 2 segmentos de codificación.

Con el valor predeterminado de 12+4, los 16 segmentos resultantes se distribuyen entre los nodos en el sitio local. Los datos y los segmentos de codificación de cada fragmento se distribuyen de manera uniforme entre los nodos del clúster. Por ejemplo, con ocho nodos, cada nodo tiene dos segmentos (de un total de 16). El motor de almacenamiento puede reconstruir un fragmento a partir de 12 de los 16 segmentos.

ECS requiere un mínimo de seis nodos para la opción de archivo inactivo, en el que se utiliza un esquema de 10+2 en lugar de 12+4. EC se detiene cuando el número de nodos está por debajo del mínimo necesario para el esquema de EC.

Cuando un fragmento está lleno o está sellado, después de un período establecido, se calcula la paridad y los segmentos de codificación se escriben en los discos en el dominio de falla. Los datos de los fragmentos permanecen como una sola copia que consta de 16 segmentos (12 datos, 4 códigos) dispersos en todo el clúster. ECS solo utiliza los segmentos de código para la reconstrucción de fragmentos cuando se produce una falla.

Cuando la infraestructura subyacente de un VDC cambia en el nivel de nodo o rack, las capas de fabric detectan el cambio y activan un escáner de rebalanceo como una tarea en segundo plano. El escáner calcula el mejor diseño para los segmentos de EC en todos los dominios de fallas de cada fragmento mediante la nueva topología. Si el nuevo diseño proporciona una mejor protección que el diseño existente, ECS vuelve a distribuir los segmentos de EC en una tarea en segundo plano. Esta tarea tiene un impacto mínimo en el rendimiento del sistema; sin embargo, habrá un aumento en el tráfico entre nodos durante el rebalanceo. También se produce el balanceo de las particiones de la tabla lógica en los nodos nuevos y los fragmentos tanto de registros como de árbol B+ recién creados se asignan uniformemente en los nodos antiguos y nuevos en el futuro. La redistribución mejora la protección local aprovechando todos los recursos dentro de la infraestructura.

Nota: se recomienda no esperar hasta que la plataforma de almacenamiento esté completamente llena antes de agregar unidades o nodos. Un umbral de utilización de almacenamiento razonable es del 70 %, teniendo en cuenta la tasa de ingesta diaria y el orden previsto, la entrega y el tiempo de integración de las unidades o los nodos agregados.

7.1 Cumplimiento de normas

A fin de cumplir con los requisitos de cumplimiento de normas corporativas y del sector (norma SEC 17a-4[f]) para el almacenamiento de datos, ECS implementó lo siguiente:

- **Reforzamiento de la plataforma**: el reforzamiento aborda las vulnerabilidades de seguridad en ECS, como el bloqueo de la plataforma para deshabilitar el acceso a los nodos o al clúster, todos los puertos no esenciales (p. ej., *ftpd, sshd*) están cerrados, el registro de auditoría completo para los comandos sudo y el soporte de SRS (Dell EMC Secure Remote services) para apagar el acceso remoto a los nodos.
- **Generación de informes de cumplimiento**: un agente del sistema que informa el estado de cumplimiento de normas del sistema como *Good*, en caso de cumplimiento, o *Bad*, en caso de falta de cumplimiento.
- **Reglas y retención de registros basada en políticas**: capacidad de limitar los cambios en los registros o los datos en retención mediante políticas, períodos y reglas.
- **Administración de retención avanzada (ARM)**: para cumplir con los requisitos de cumplimiento de normas de Centera, se definió un conjunto de reglas de retención solo para CAS.
  - **Retención basada en eventos**: permite periodos de retención que comienzan cuando ocurre el evento especificado.
  - **Retención para asuntos legales y de auditoría**: permite la prevención de la eliminación temporal de datos sujeta a acciones legales.
  - **Controladora mín./máx.**: configuración por depósito para los períodos de retención mínimo y máximo predeterminados.

El cumplimiento de normas está activado en el nivel del espacio de nombres. Los períodos de retención se configuran en el nivel del depósito. Los requisitos de cumplimiento de normas certifican la plataforma y, debido a esto, la función de cumplimiento de normas solo está disponible para ECS que se ejecuta en el hardware del dispositivo. Para obtener información sobre cómo habilitar y configurar el cumplimiento de normas en ECS, consulte la *Guía de acceso a datos de ECS* actual y la *Guía del administrador de ECS* más reciente.

8 Implementación

ECS se puede implementar como una instancia de un solo sitio o de múltiples sitios. Los elementos esenciales de una implementación de ECS incluyen lo siguiente:

- **Centro de datos virtual (VDC)**: un clúster, también conocido generalmente como un sitio o una región geográficamente distinta, compuesto por un conjunto de infraestructura de ECS administrado por una sola instancia de fabric.
- **Pool de almacenamiento (SP)**: los SP se pueden considerar como un subconjunto de nodos y que su almacenamiento asociado pertenece a un VDC. Un nodo puede pertenecer a un solo SP. EC está configurado en el nivel de SP con un esquema de 12+4 o 10+2. Un SP se puede usar como una herramienta para separar físicamente los datos entre los clientes o grupos de clientes que acceden al almacenamiento en ECS.
- **Grupo de replicación (RG)**: el grupo de replicación define dónde se protege el contenido del SP y las ubicaciones desde las cuales se puede acceder a los datos. A veces, un grupo RG con un solo sitio miembro se denomina RG local. Los datos siempre se protegen de manera local, donde se escriben contra fallas de disco, de nodo y de rack. Los RG con dos o más sitios suelen denominarse RG globales. Los RG globales abarcan hasta 8 VDC y protección contra fallas de discos, nodos, racks y sitios. Un VDC puede pertenecer a múltiples RG.
- **Espacio de nombres**: un espacio de nombres es conceptualmente lo mismo que un grupo de usuarios en ECS. Una característica clave de un espacio de nombres es que los usuarios de un espacio de nombres no pueden acceder a los objetos que pertenezcan a otro espacio de nombres.
- **Depósitos**: los depósitos son contenedores de objetos creados en un espacio de nombres y, a veces, se consideran un contenedor lógico para los subgrupos de usuarios. En S3, los contenedores se denominan depósitos y ECS ha adoptado este término. En Atmos, el equivalente a un depósito es un subusuario; en Swift, el equivalente de un depósito es un contenedor, y, en CAS, un depósito es un pool de CAS. Los depósitos son recursos globales de ECS. Cada depósito se crea en un espacio de nombres y cada espacio de nombres se crea en un RG.

ECS aprovecha los siguientes sistemas de infraestructura:

- **DNS**: (necesario) búsquedas directas e inversas necesarias para cada nodo de ECS.
- **NTP**: (necesario) servidor de Network Time Protocol.
- **SMTP**: (opcional) Simple Mail Transfer Protocol Server para el envío de alertas y la generación de informes.
- **DHCP**: (opcional) es necesario si se asignan direcciones IP a través de DHCP.
- **Proveedores de autenticación**: (opcional) los administradores de ECS pueden autenticarse mediante grupos de LDAP y Active Directory. Los usuarios de objetos pueden autenticarse mediante Keystone. Los proveedores de autenticación no son necesarios para ECS. ECS tiene incorporada la funcionalidad de administración de usuarios locales; sin embargo, tenga en cuenta que los usuarios creados localmente no se replican entre VDC.
- **Balanceador de carga**: (necesario si el flujo de trabajo lo determina, de lo contrario, es opcional) la carga del cliente debe distribuirse entre los nodos para utilizar de manera eficaz todos los recursos disponibles en el sistema. Si se necesita un servicio o un dispositivo balanceador de carga dedicado para administrar la carga entre los nodos de ECS, debe considerarse como un requisito. Los desarrolladores que escriben aplicaciones mediante el SDK de S3 de ECS pueden aprovechar su funcionalidad de balanceador de carga integrado. Los balanceadores de carga sofisticados pueden incluir factores adicionales, como la carga informada, los tiempos de respuesta, el estado activo/inactivo, el número de conexiones activas
y la ubicación geográfica del servidor. El cliente es responsable de administrar el tráfico del cliente y de determinar los requisitos de acceso. Independientemente del método, hay algunas opciones básicas que se consideran generalmente e incluyen la asignación manual de direcciones IP, la Round Robin de DNS, el balanceo de carga en el lado del cliente, los dispositivos balanceadores de carga y los balanceadores de carga geográficos. A continuación se presentan breves descripciones de cada uno de estos métodos:

- **Asignación manual de IP**: las direcciones IP se distribuyen manualmente a las aplicaciones. Por lo general, esto no se recomienda porque es posible que no distribuya la carga ni proporcione tolerancia a fallas.
- **Round Robin de DNS**: se crea una entrada de DNS que incluye todas las direcciones IP del nodo. Los clientes consultan DNS para resolver nombres de dominio completamente calificados para los servicios de ECS y se responden con las direcciones IP de un nodo aleatorio. Esto puede proporcionar un seudo balanceo de carga. Es posible que este método no proporcione tolerancia a fallas debido a que, a menudo, se utiliza la intervención manual para eliminar direcciones IP de nodos fallidos de DNS. Con este método se pueden encontrar los problemas de tiempo de disponibilidad (TTL). Algunas implementaciones de servidores DNS pueden almacenar en caché las búsquedas de DNS durante un período de tiempo, de modo que los clientes que se conectan en un plazo de cierre puedan vincularse con la misma dirección IP, lo que reduce la cantidad de distribución de la carga a los nodos de datos. No se recomienda el uso de DNS para la distribución de tráfico en modo Round Robin.
- **Balanceo de carga**: los balanceadores de carga son el enfoque más común para distribuir la carga del cliente. Los clientes pueden enviar el tráfico a un balanceador de carga que lo recibe y reenvía a un nodo de ECS en buen estado. El estado de conexión o las evaluaciones del estado proactivas se usan para verificar la disponibilidad de cada nodo para las solicitudes de servicio. Los nodos no disponibles no se pueden utilizar hasta que pasen una evaluación del estado. La descarga del procesamiento SSL intensivo de CPU se puede usar para liberar esos recursos en ECS.
- **Balanceo de carga geográfico**: el balanceo de carga geográfico aprovecha el DNS para enrutar búsquedas a un dispositivo, como Riverbed SteelApp, por ejemplo, que utiliza Geo-IP u otro mecanismo para determinar el mejor sitio al cual dirigir al cliente.

8.1 Implementación en un solo sitio

Durante una implementación inicial de un solo sitio o de un solo clúster, los nodos son los primeros en agregarse a un SP. Los SP son contenedores lógicos de nodos físicos. La configuración del SP implica la selección del número mínimo necesario de nodos disponibles y la selección del esquema de EC de 12+4 predeterminado o de 10+2 de archivo inactivo. Los niveles de alerta críticos pueden configurarse inicialmente durante la configuración del SP y en el futuro; sin embargo, el esquema de EC no se puede cambiar después de la inicialización del SP. El primer SP creado se designa como el SP del sistema y se utiliza para almacenar los metadatos del sistema. El SP del sistema no se puede eliminar.

Por lo general, los clústeres contienen uno o dos SP, como se muestra en la Figura 21, uno para cada esquema de EC; sin embargo, si una organización necesita de la separación física de los datos, los SP adicionales se utilizan para implementar límites.

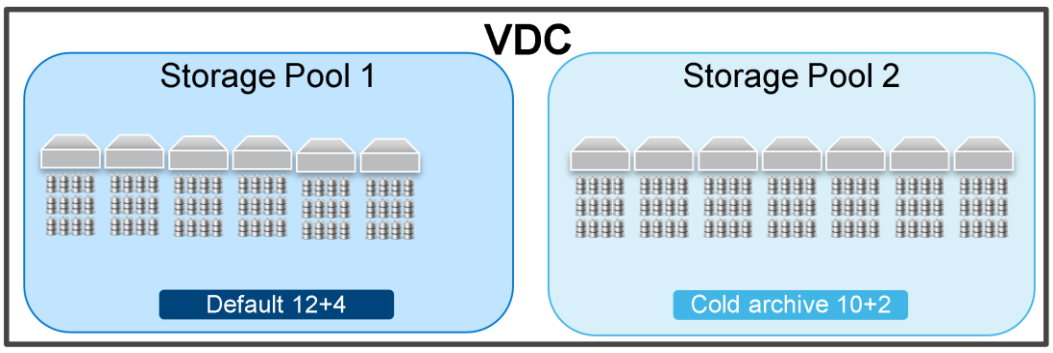

Figura 21 VDC con dos pools de almacenamiento, cada uno configurado con un esquema de EC diferente

Después de la inicialización del primer SP, se puede crear un VDC. La configuración de VDC implica la designación de terminales de administración y replicación. Tenga en cuenta que, a pesar de que es necesaria la inicialización del SP del sistema antes de la creación del VDC, la configuración de VDC no asigna los SP, sino las direcciones IP de los nodos.

Después de que se crea un VDC, se configuran los RG. Los RG son recursos globales con una configuración que implica la designación de al menos un VDC en la configuración de un solo sitio o de un solo sitio inicial, junto con uno de los SP del VDC. Un RG con un solo miembro de VDC protege los datos de manera local en el nivel de disco, nodo y rack. En la siguiente sección se explican más los RG para incluir implementaciones de varios sitios.

Los espacios de nombres son recursos globales creados y asignados a un RG. En las políticas de retención de nivel de espacio de nombres, se definen las cuotas, el cumplimiento de normas y los administradores de espacio de nombres. El acceso durante la interrupción (ADO) puede configurarse en el nivel de espacio de nombres, que se aborda en la sección siguiente. Por lo general, es en el nivel de espacio de nombres donde se organizan los grupos de usuarios. Los grupos de usuarios pueden ser una instancia de la aplicación o un equipo, un usuario, un grupo de empresas o cualquier otra agrupación que tenga sentido para la organización.

Los depósitos son recursos globales y pueden abarcar varios sitios. La creación de depósitos implica asignarlo a un espacio de nombres y a un RG. El nivel de depósito es donde se habilita la propiedad y el acceso a archivos o CAS. La Figura 22 que aparece a continuación muestra un SP en un VDC con un espacio de nombres que contiene dos depósitos.

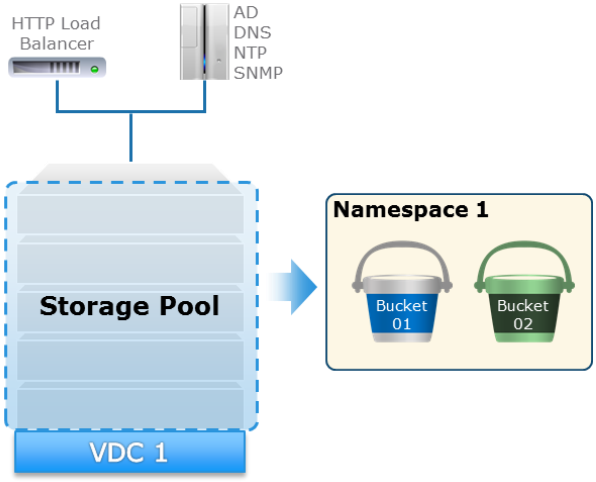

Figura 22 Ejemplo de implementación de un solo sitio

8.2 Implementación de múltiples sitios

Una implementación de varios sitios, también conocida como un entorno federado o ECS federado, puede abarcar hasta ocho VDC. Los datos se replican en ECS en el nivel de fragmento. Los nodos que participan en un RG envían sus datos locales de manera asíncrona a uno o a todos los demás sitios. Los datos se cifran mediante AES256 antes de que se envíen a través de la WAN mediante HTTP. Los beneficios clave reconocidos cuando se crea una federación de varios VDC son los siguientes:

- Consolidación de esfuerzos de administración de varios VDC en un único recurso lógico
- Protección en el nivel de sitio, además de localmente, para los niveles de nodo, disco y rack
- Acceso distribuido geográficamente para almacenamiento de manera totalmente coherente y activa en todas partes

En esta sección sobre implementación de varios sitios se describen las funciones específicas de ECS federado, como las siguientes:

- **Coherencia de datos**: de manera predeterminada, ECS ofrece un servicio de almacenamiento totalmente coherente.
- **Grupos de replicación**: contenedores globales que se utilizan para designar límites de acceso y protección.
- **Almacenamiento en caché geográfico**: optimización de los flujos de trabajo de acceso a sitios remotos en implementaciones de múltiples sitios.
- **ADO**: comportamiento de acceso de cliente durante un corte de suministro eléctrico temporal en el sitio (TSO).

8.2.1 Coherencia de datos

ECS es un sistema totalmente coherente que utiliza la propiedad para mantener una versión autorizada de cada espacio de nombres, depósito y objeto. La propiedad se asigna al VDC donde se crea el espacio de nombres, el depósito o el objeto. Por ejemplo, si se crea un espacio de nombres, NS1, en VDC1, entonces VDC1 es propietario de NS1 y es responsable de mantener la versión autorizada de los depósitos dentro de NS1. Si se crea un depósito, B1, en VDC2 dentro de NS1, entonces VDC2 es el propietario de B1 y es responsable de mantener la versión autorizada de los contenidos del depósito, así como el VDC propietario de cada objeto. De manera similar, si un objeto, O1, se crea dentro de B1 en VDC3, entonces VDC3 es el propietario de O1 y es responsable de mantener la versión autorizada de O1 y los metadatos asociados.

La resiliencia de la protección de datos de varios sitios conlleva el costo de una mayor sobrecarga de protección de almacenamiento y el consumo de ancho de banda de WAN. Las consultas de índice son necesarias cuando se accede a un objeto o se actualiza desde un sitio que no es el propietario del objeto. De manera similar, las búsquedas de índice a través de la WAN también son necesarias para recuperar información como una lista autorizada de depósitos en un espacio de nombres u objetos en un depósito, que son propiedad de un sitio remoto.

Comprender cómo ECS utiliza la propiedad para rastrear los datos de manera autorizada en el nivel de espacio de nombres, depósito y objetos ayuda a los administradores y a los propietarios de aplicaciones a tomar decisiones sobre la configuración de su entorno para el acceso.

8.2.2 Grupo de replicación activo

Durante la creación de un RG, el ajuste para *Replicar a todos los sitios* está disponible, el cual queda desactivado, de forma predeterminada, o se puede alternar a activado para habilitar esta función.
La replicación de datos a todos los sitios significa que los datos escritos individualmente en cada VDC se replican al resto de VDC miembros del RG. Por ejemplo, una instancia de ECS federada de X cantidad de sitios que cuenta con un RG activo configurado para replicar datos a todos los sitios generará una sobrecarga de protección de X veces, o sobrecarga total de protección de datos de X * 1,33 (o 1,2 en el EC de archivo inactivo). La replicación a todos los sitios puede tener sentido especialmente para conjuntos de datos más pequeños donde el acceso local es importante. Dejar este ajuste desactivado significa que todos los datos escritos en cada VDC se replicarán en otro VDC. El sitio principal, donde se crea el objeto, y el sitio que almacena la copia de replicación, cada uno protege los datos de manera local mediante el esquema de EC asignado al SP local. Es decir, que solo los datos originales se replican a través de la WAN y no a los segmentos de codificación de EC asociados.

Los clientes pueden acceder a los datos almacenados en un RG activo a través de cualquier VDC miembro de RG disponible. La Figura 23 que aparece a continuación muestra un ejemplo de un ECS federado desarrollado con VDC1, VDC2 y VDC3. Se muestran dos RG: RG1 tiene un solo miembro, VDC1, y el RG2 tiene tres VDC como miembros. Se muestran tres depósitos: B1, B2 y B3.

En este ejemplo, los clientes acceden a lo siguiente:

- VDC1 tiene acceso a todos los depósitos
- VDC2 y VDC3 tienen acceso solo a los depósitos B2 y B3.

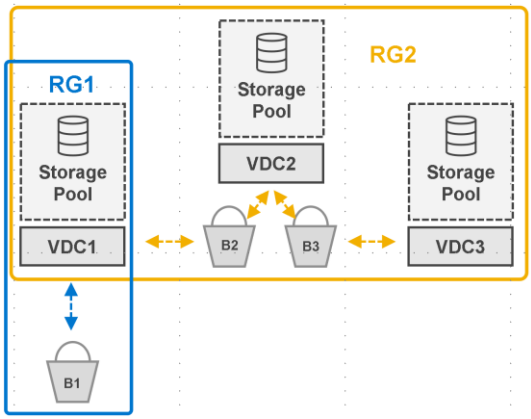

Figura 23 Acceso en el nivel de depósito por sitio con grupos de replicación de un solo sitio y de múltiples sitios

8.2.3 Grupo de replicación pasivo

Un RG pasivo tiene tres VDC miembros. Dos de los VDC se designan como activos y son accesibles para los clientes. El tercer VDC se designa como pasivo y se usa como objetivo de replicación solamente. El sitio pasivo se utiliza solo con fines de recuperación y no permite el acceso directo del cliente. Los beneficios de la replicación geopasiva son los siguientes:

- Reducción en la sobrecarga de la protección del almacenamiento mediante el aumento del potencial de las operaciones XOR
- Control en el nivel de administrador de la ubicación utilizada para el almacenamiento solo de replicación

Figura 24 muestra un ejemplo de una configuración geopasiva en la cual el VDC 1 y el VDC 2 son sitios principales (fuente) que replican sus datos (fragmentos) en el objetivo de replicación, VDC 3.

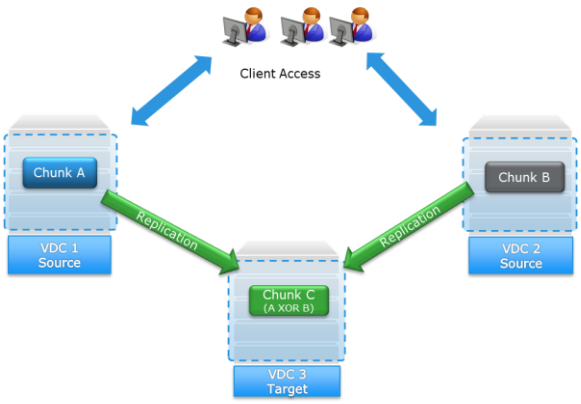

Figura 24 Rutas de acceso de cliente y replicación para el grupo de replicación geopasiva

El acceso de varios sitios a datos totalmente coherentes se logra mediante el uso de la propiedad de objetos, espacio de nombres y depósito en los sitios miembros de RG. Las consultas de indexación entre sitios a través de la WAN son necesarias cuando el acceso a la API se origina en un VDC que no es el propietario de las construcciones lógicas necesarias. Las búsquedas de WAN se utilizan para determinar la versión autorizada de los datos. Por lo tanto, si un objeto creado en el sitio 1 se lee desde el sitio 2, es necesaria una búsqueda de WAN para consultar el VDC propietario del objeto, el sitio 1, para verificar si los datos del objeto que se replicaron en el sitio 2 son la versión más reciente de los datos. Si el sitio 2 no tiene la versión más reciente, obtiene los datos necesarios del sitio 1, de lo contrario, utiliza los datos que se replicaron anteriormente hacia él. Esto se muestra en la Figura 25 que aparece a continuación.

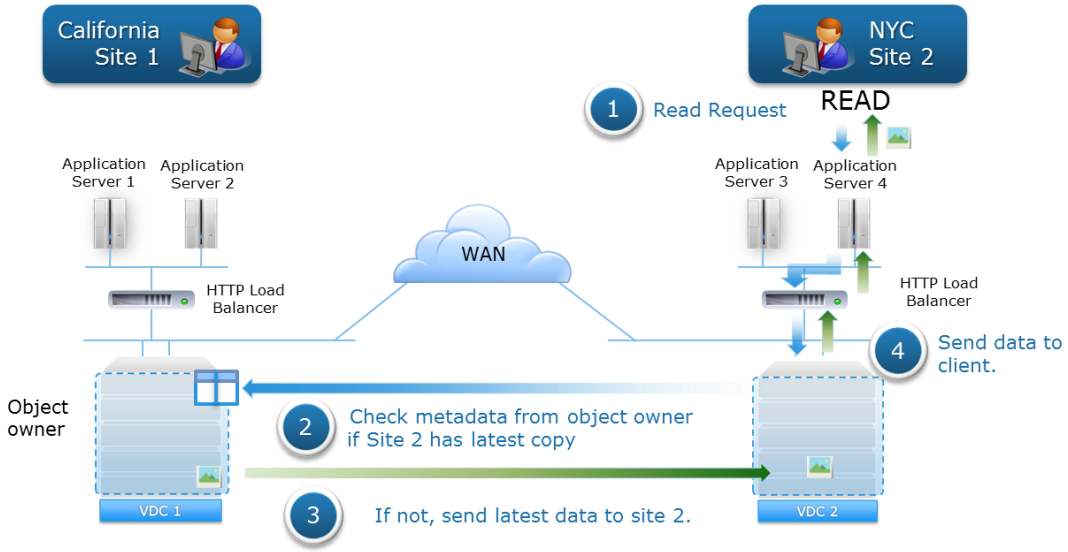

Figura 25 La solicitud de lectura a un VDC que no es propietario activa la búsqueda de WAN en el VDC propietario del objeto

El flujo de datos de escrituras en un entorno replicado geográficamente en el que dos sitios están actualizando el mismo objeto se muestra en la Figura 26. En este ejemplo, el sitio 1 se crea inicialmente y es propietario del objeto. El objeto cuenta con codificación de eliminación y las transacciones de registro relacionadas se escriben en el disco en el sitio 1. El flujo de datos para una actualización del objeto recibido en el sitio 2 es el siguiente:

1. En primer lugar, el sitio 2 escribe los datos localmente.
2. El sitio 2 actualiza de manera síncrona los metadatos (escritura de registro) con el propietario del objeto, el sitio 1, y espera la confirmación de la actualización de metadatos del sitio 1.

3. El sitio 1 confirma la escritura de metadatos en el sitio 2.
4. El sitio 2 confirma la escritura en el cliente.

Nota: el sitio 2 replica de manera asíncrona el sitio de datos 1, el sitio propietario del objeto, con normalidad. Si los datos se deben obtener desde el sitio 1 antes de que se repliquen desde el sitio 2, el sitio 1 recuperará los datos directamente desde el sitio 2.

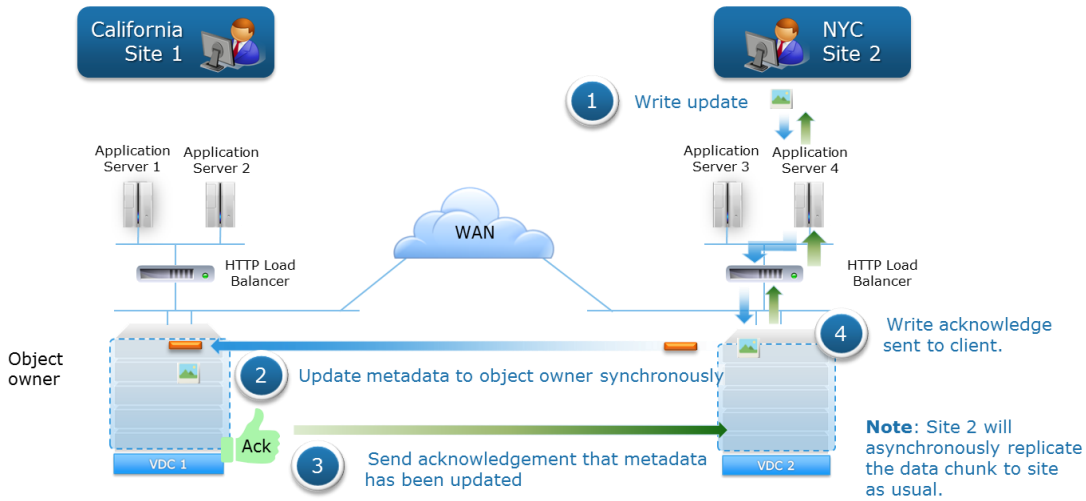

Figura 26 Actualización del mismo flujo de datos de objetos en un entorno replicado geográficamente

En escenarios de lectura y escritura en un entorno replicado geográficamente, hay latencia involucrada en la lectura y la actualización de los metadatos, y la recuperación de datos desde el sitio propietario de objetos.

Nota: desde ECS 3.4, puede quitar un VDC de un grupo de replicación (RG) de una federación con diversos VDC sin afectar al VCD u otros RG asociados con él. La eliminación del VDC de un RG ya no inicia la PSO (corte de suministro eléctrico del sitio permanente). La eliminación de un VDC del RG inicia la recuperación.

Consulte la *Guía del administrador de ECS* más reciente para obtener más información sobre el grupo de replicación

8.2.4 Almacenamiento en caché geográfico de datos remotos

ECS optimiza los tiempos de respuesta para obtener acceso a los datos almacenados en sitios remotos mediante el almacenamiento en caché local de los objetos que se leen a través de la WAN. Esto puede ser útil para los patrones de acceso de múltiples sitios en los que los datos a menudo se recuperan desde un sitio remoto o que no es el propietario. Considere un entorno replicado geográficamente con tres sitios, VDC1, VDC2 y VDC3, donde se escribe un objeto en VDC1 y la copia de replicación del objeto se almacena en VDC2. En este escenario, para el servicio de una solicitud de lectura recibida en VDC3 en cuanto a un objeto que se creó en VDC1 y se replicó a VDC2, los datos de objetos se deben enviar a VDC3 desde VDC1 o VDC2. Los datos remotos almacenados en caché geográfico a los que se accede con frecuencia ayudan a reducir los tiempos de respuesta. Se emplea un algoritmo menos utilizado recientemente para el almacenamiento en caché. El tamaño de la caché geográfica se ajusta cuando la infraestructura de hardware, como discos, nodos y racks, se agregan a un SP replicado geográficamente.

8.2.5 Comportamiento durante un corte de suministro eléctrico en el sitio

Un corte de suministro eléctrico temporal sitio (TSO) generalmente se refiere a una falla de conectividad WAN o de un sitio completo, por ejemplo, durante un desastre natural. ECS utiliza mecanismos de latido para detectar y manejar fallas del sitio temporales. El acceso de clientes y la disponibilidad de las operaciones de la API en los niveles de espacio de nombres, depósito y objetos durante un TSO se rigen por las siguientes opciones de ADO configuradas en el nivel de espacio de nombres y depósito:

- **Desactivado (predeterminado)**: se mantiene una alta coherencia durante un corte de suministro eléctrico temporal.
- **Activado**: durante un corte de suministro eléctrico temporal en el sitio, finalmente se permite el acceso coherente.

La coherencia de datos durante un TSO se implementa en el nivel del depósito. La configuración se establece en el nivel del espacio de nombres, lo que establece la configuración de ADO predeterminada para ADO durante la creación de un depósito nuevo. y se pueden reemplazar en la creación de depósitos nuevos; lo que significa que TSO se puede configurar para algunos depósitos y no para otros.

8.2.5.1 Acceso durante corte de suministro eléctrico (ADO) no activado

De forma predeterminada, ADO no está activado y se mantiene una coherencia alta. Todas las solicitudes de API de cliente en las que se necesiten datos autorizados de espacio de nombres, depósitos u objetos, pero que no están disponibles temporalmente, fallarán. Las operaciones de objetos de lectura, creación, actualización y eliminación, así como depósitos de lista que no pertenecen a un sitio en línea, fallarán. Además, las operaciones de creación y edición de depósito, usuario y espacio de nombres también fallarán.

Como se mencionó anteriormente, el sitio inicial propietario del depósito, espacio de nombres y objeto es el sitio donde se creó el recurso por primera vez. Durante un TSO, ciertas operaciones pueden fallar si no se puede acceder al sitio propietario del recurso. Los aspectos importantes de las operaciones permitidas o no permitidas durante un corte de suministro eléctrico del sitio incluyen los siguientes:

- No se permite la creación, la eliminación ni la actualización de depósitos, espacios de nombres, usuarios de objetos, proveedores de autenticación, usuarios de NFS y RG, y el mapeo de grupos desde ningún sitio.
- Se permite enumerar depósitos dentro de un espacio de nombres si el sitio propietario del espacio de nombres está disponible.

HDFS/NFS habilita que los depósitos que son propiedad del sitio inaccesible sean de solo lectura.

8.2.5.2 ADO habilitado

En un depósito habilitado para ADO, durante un TSO, el servicio de almacenamiento proporciona respuestas coherentes finalmente. En este escenario, las lecturas y, de manera opcional, las escrituras de un sitio secundario (no propietario) se aceptan y se respetan. Además, una escritura en un sitio secundario durante un TSO hace que este sitio tome la propiedad del objeto. Esto permite que cada VDC siga leyendo y escribiendo objetos de depósitos en un espacio de nombres compartido. Por último, la nueva versión del objeto se convierte en la versión autorizada del objeto durante la conciliación después del TSO, incluso si otra aplicación actualiza el objeto en el VDC propietario.

A pesar de que muchas operaciones de objetos continúan durante un corte de suministro eléctrico en la red, no se permiten ciertas operaciones, como la creación de nuevos depósitos, espacios de nombres o usuarios. Cuando se restaura la conectividad de red entre dos VDC, el mecanismo de latido detecta automáticamente la conectividad, restaura el servicio y reconcilia los objetos de los dos sitios. Si se actualiza el mismo objeto en el VDC A y el VDC B, la copia en el VDC que no es propietario es la copia autorizada. Por lo tanto, si un objeto que es propiedad de VDC B se actualiza en el VDC A y el VDC B durante la sincronización, la copia

en el VDC A será la copia autorizada que se mantiene, y no se podrá hacer referencia a la otra copia y estará disponible para la recuperación de espacio.

Cuando más de dos VDC forman parte de un RG, y si se interrumpe la conectividad de red entre un VDC y los otros dos, entonces las operaciones de escritura/actualización/propiedad continúan tal como lo harían con dos VDC, pero el proceso para responder a solicitudes de lectura es más complejo, como se describe a continuación.

Si una aplicación solicita un objeto que es propiedad de un VDC que no está accesible, ECS envía la solicitud al VDC con la copia secundaria del objeto. Sin embargo, la copia del sitio secundaria puede haber estado sujeta a una operación de contracción de datos, que es una operación XOR entre dos conjuntos de datos diferentes que produce un nuevo conjunto de datos. Por lo tanto, el VDC del sitio secundario primero debe recuperar los fragmentos del objeto que se incluyen en la operación XOR original y debe haber realizado una operación XOR en esos fragmentos con la copia de recuperación. Esta operación devuelve el contenido del fragmento almacenado originalmente en el VDC de falla. Luego, los fragmentos del objeto recuperado se pueden reconstruir y devolver. Cuando se reconstruyen los fragmentos, también se almacenan en caché para que el VDC pueda responder más rápidamente a las solicitudes subsiguientes. Tenga en cuenta que la reconstrucción lleva mucho tiempo. Una mayor cantidad de VDC en un RG implica una mayor cantidad de fragmentos que se deben recuperar de otros VDC y, por lo tanto, la reconstrucción del objeto tarda más tiempo.

Si se produce un desastre, el VDC completo puede ser irrecuperable. ECS trata el VDC irrecuperable como una falla temporal del sitio. Si la falla es permanente, el administrador del sistema debe realizar una conmutación por error permanentemente del VDC desde la federación para iniciar el procesamiento de conmutación por error; esto inicia la resincronización y vuelve a proteger los objetos almacenados en el VDC fallido. Las tareas de recuperación se ejecutan como un proceso en segundo plano. Puede revisar el progreso de la recuperación en el portal de ECS.

Una opción de depósito adicional está disponible para ADO de *solo lectura (RO)*, lo que garantiza que la propiedad de los objetos nunca cambie y elimina la posibilidad de que se produzcan conflictos causados por las actualizaciones de objetos en los sitios fallidos y en línea durante un corte de suministro eléctrico temporal en el sitio. La desventaja de ADO de RO es que, durante un corte de suministro eléctrico temporal en el sitio, no se pueden crear objetos nuevos ni se pueden actualizar objetos existentes en el depósito hasta que todos los sitios estén nuevamente en línea. La opción ADO de RO está disponible solamente durante la creación del depósito; no se puede modificar posteriormente. De forma predeterminada, esta opción está desactivada.

Tabla 8 Tolerancia a fallas de múltiples sitios

Modelo de falla	Tolerancia
Entorno replicado geográficamente	Hasta una falla del sitio

8.3 Tolerancia a fallas

ECS está diseñado para tolerar una variedad de situaciones de falla de equipos mediante varios dominios de fallas. El rango de condiciones de falla abarca un alcance variable, que incluye lo siguiente:

- Falla de un solo disco duro en un nodo único
- Falla de varios discos duros en un nodo único
- Varios nodos con falla de un solo disco duro

- Varios nodos con varias fallas de disco duro
- Falla de un nodo único
- Falla de varios nodos
- Pérdida de comunicación con un VDC replicado
- Pérdida de un VDC replicado completo

En una configuración de un solo sitio, de dos sitios o de replicación geográfica, el impacto de la falla depende de la cantidad y el tipo de componentes afectados. Sin embargo, en cada nivel, ECS proporciona mecanismos de defensa contra el impacto de las fallas de componentes. Muchos de estos mecanismos ya se analizaron en este informe, pero se analizan aquí y en la Figura 27 para mostrar cómo se aplican a la solución. Estas incluyen:

- Disk failure
  - Los segmentos de EC o las copias de réplica del mismo fragmento no se almacenan en el mismo disco
  - Cálculo de suma de comprobación en operaciones de lectura y escritura
  - Comprobador de coherencia en segundo plano que vuelve a verificar las sumas de comprobación
- Node failure
  - Distribuir fragmentos o copias de réplica de un fragmento por igual entre los nodos de un VDC
  - Fabric de ECS mantiene los servicios en ejecución y administra los recursos, como los discos y la red.
  - Las tablas y los registros de partición protegidos por conmutación por error de propiedad de la partición de nodo a nodo.
- Falla del rack en VDC
  - Distribuir segmentos de copias de réplica de un fragmento por igual entre los racks de un VDC.
  - Una instancia de registro de fabric se ejecuta en cada rack y se puede reiniciar en cualquier otro nodo en el mismo rack si el nodo falla.

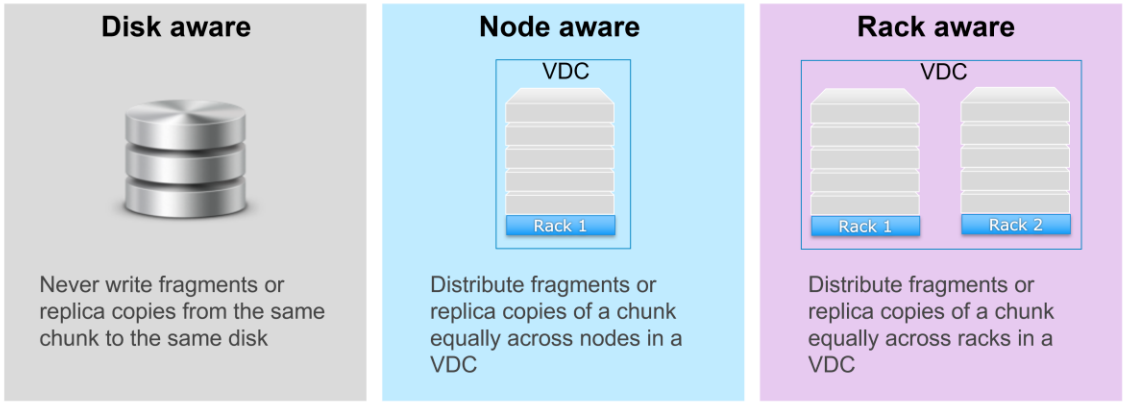

Figura 27 Mecanismos de protección en los niveles de disco, nodo y rack

En la siguiente tabla se define el tipo y el número de fallas de componentes que cada esquema de EC protege por configuración de rack básico. La Tabla 9 destaca la importancia de considerar el impacto de los dominios de fallas de protección en la disponibilidad general de datos y servicios en términos del número de nodos necesarios en cada esquema de EC.

Tabla 9 Protección del código de eliminación en los dominios de fallas

Esquema de EC	N.º de nodos en VDC	N.º de partes de fragmento por nodo	Datos de EC protegidos contra…
12+4 Predeterminado	5 o menos	4	• Pérdida de hasta cuatro discos o • Pérdida de un nodo
	6 u 7	3	• Pérdida de hasta cuatro discos o • Pérdida de un nodo y de un disco de un segundo nodo
	8 o más	2	• Pérdida de hasta cuatro discos o • Pérdida de dos nodos o • Pérdida de un nodo y de dos discos
	16 o más	1	• Pérdida de cuatro nodos o • Pérdida de tres nodos y discos de un nodo adicional o • Pérdida de dos nodos y discos de hasta dos nodos diferentes o • Pérdida de un nodo y de discos de hasta tres nodos diferentes o • Pérdida de cuatro discos de cuatro nodos diferentes
10+2 Almacenamiento inactivo	11 o menos	2	• Pérdida de hasta dos discos o • Pérdida de un nodo
	12 o más	1	• Pérdida de cualquier cantidad de discos de dos nodos diferentes o • Pérdida de dos nodos

8.4 Automatización de reemplazo de disco

A partir de ECS 3.5, los clientes pueden reemplazar los discos fallidos por los servicios de Dell EMC mediante un flujo de trabajo intuitivo del portal de ECS (interfaz de usuario web). La función proporciona lo siguiente:

- Resolución de hágalo usted mismo para los errores de unidad
- Tiempo acelerado para la reparación de fallas
- Flexibilidad operacional y ahorros de TCO

La página de mantenimiento en el portal de ECS proporciona visibilidad de administrador para todos los discos de cada nodo. Cuando una unidad falla, el sistema inicia automáticamente la recuperación. Se recuperan todos los tipos de recursos en la unidad y, cuando la unidad esté lista para eliminarse del nodo, el portal de ECS mostrará el botón reemplazar, como se muestra en Figura 28.

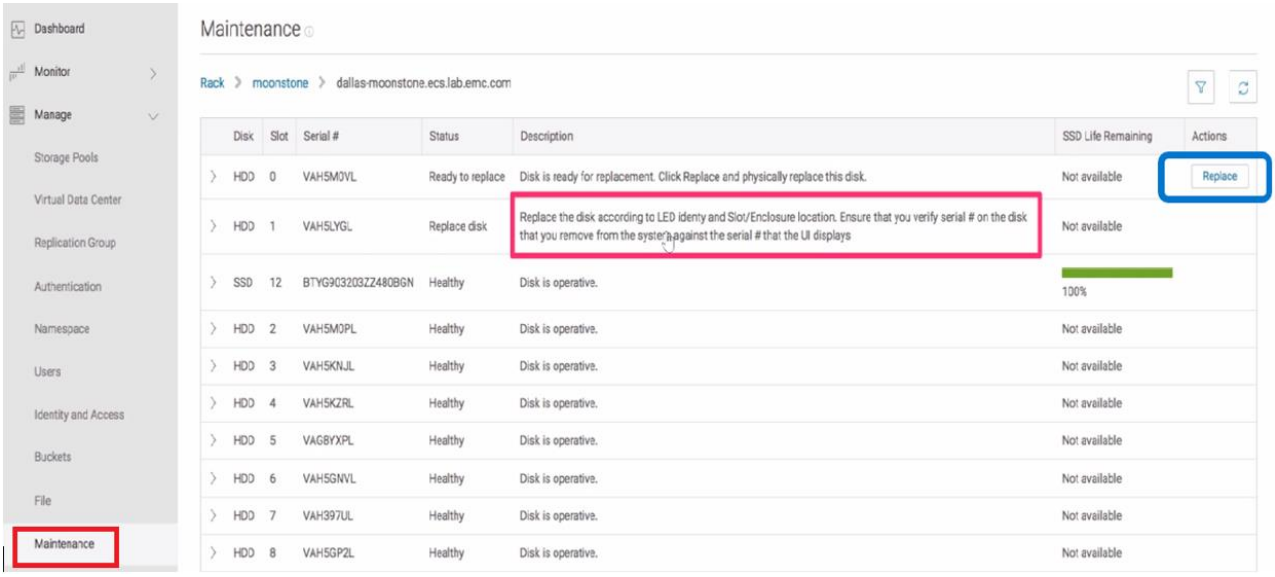

Figura 28 Automatización de reemplazo de discos

Nota: se debe reemplazar solo una unidad por vez. Esto es para evitar el reemplazo de la unidad incorrecta.

8.5 Tech Refresh

Tech Refresh es una contratación dirigida por Dell EMC Professional Services disponible a partir de ECS 3.5 para eliminar de manera no disruptiva los nodos de hardware más antiguos de los clústeres de ECS mediante la función de software integrada. Es una operación eficiente y de bajo consumo de recursos que se puede regular con precisión. Esta función reduce la sobrecarga asociada anteriormente con la desactivación del hardware de ECS.

La actualización tecnológica incluye tres partes:

- **Extensión de nodos**: adición de nodos de 3.ª generación al clúster existente
- **Migración de recursos**: transfiera todos los recursos de los nodos existentes a los nodos de 3.ª generación
- **Evacuación de nodos**: limpie los nodos antiguos y quítelos del clúster

Los servicios profesionales deben participar durante el mantenimiento de la actualización tecnológica. Consulte la *Guía de actualización tecnológica de ECS* más reciente para obtener más información sobre la actualización tecnológica.

9 Sobrecarga de protección de almacenamiento

Cada miembro de VDC de un RG es responsable de su propia protección de datos de EC en el nivel local. Es decir, los datos se replican, pero no cualquier segmento de codificación relacionado. A pesar de que EC tiene una mayor eficiencia del almacenamiento que otras formas de protección, como el espejeado de unidades de copia completa, implica una sobrecarga de costos de almacenamiento inherente en el nivel local. Sin embargo, cuando es necesario tener copias secundarias replicadas fuera del sitio, al igual que todos los sitios tengan acceso a los datos cuando un solo sitio deja de estar disponible, los costos de almacenamiento aumentan más que cuando se utilizan los métodos tradicionales de protección de copia de datos de sitio a sitio. Esto es especialmente cierto cuando los datos únicos se distribuyen entre tres o más sitios.

ECS proporciona un mecanismo en el cual la eficiencia de la sobrecarga de protección de almacenamiento puede aumentar a medida que se federan tres o más sitios. En un entorno replicado de dos VDC, ECS replica fragmentos desde el VDC primario, o el propietario, a un sitio remoto para proporcionar alta disponibilidad y resiliencia. No hay manera de eludir el costo del 100 % de la sobrecarga de protección de una copia completa de los datos en una implementación de ECS federada de dos sitios.

Ahora, considere tres VDC en un entorno de múltiples sitios, VDC1, VDC2 y VDC3, donde cada VDC tiene datos únicos que se replican a ellos desde cada uno de los otros VDC. VDC2 y VDC3 pueden enviar una copia de sus datos a VDC1 para la protección. Por lo tanto, VDC1 tendría sus propios datos originales, además de replicar los datos de VDC2 y VDC3. Esto significa que VDC1 almacenará 3 veces la cantidad de datos escritos en su propio sitio.

En esta situación, ECS puede realizar una operación XOR de datos de VDC2 y VDC3 almacenados localmente en VDC1. Esta operación matemática compara la misma cantidad de datos únicos y fragmentos, y genera un resultado en un nuevo fragmento que contiene suficientes características de los dos fragmentos de datos originales para permitir la restauración de cualquiera de los dos conjuntos originales. Por lo tanto, en los casos en los que antes había tres conjuntos únicos de fragmentos de datos almacenados en VDC1, consumiendo 3 veces la capacidad disponible, ahora solo hay dos de los conjuntos de datos locales originales y las copias de protección de XOR reducidas.

En este mismo caso, si VDC3 deja de estar disponible, ECS puede reconstruir los fragmentos de datos de VDC3 usando copias de fragmentos recuperadas de VDC2 y los datos (C1 ⊕ C2) de VDC3 almacenados localmente en VDC1. Este principio se aplica a los tres sitios que participan en el RG y depende de que cada uno de los tres VDC tenga conjuntos de datos únicos. La Figura 29 muestra un cálculo XOR con dos sitios que se replican en un tercer sitio.

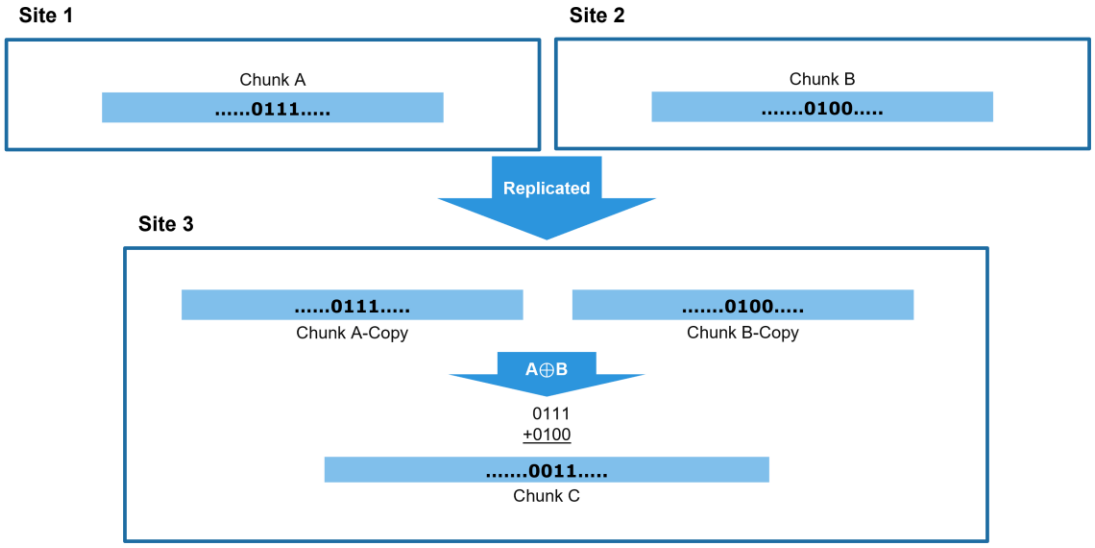

Figura 29 Eficiencia de la protección de datos XOR

Si los acuerdos de nivel de servicio de la empresa requieren velocidades de acceso de lectura óptimas incluso en caso de una falla del sitio completo, la configuración de replicación a todos los sitios obliga a ECS a revertir a copias completas de los datos replicados que se almacenarán en todos los sitios. Como es de esperar, esto impulsa los costos de almacenamiento en proporción a la cantidad de VDC que participan en el RG. Por lo tanto, una configuración de 3 sitios volvería a tres veces la sobrecarga de protección de almacenamiento. La configuración Replicar a todos los sitios está disponible durante la creación del RG y no se puede alternar entre sí.

A medida que aumenta la cantidad de sitios federados, la optimización de XOR es más eficiente en la reducción de la sobrecarga de protección de almacenamiento debido a la replicación. La Tabla 10 proporciona información sobre la sobrecarga de protección de almacenamiento en función del número de sitios para EC normal de 12+4 y EC de archivo inactivo de 10+2, lo que ilustra cómo ECS puede tener mayor eficiencia del almacenamiento a medida que se enlazan más sitios.

Nota: para reducir la sobrecarga de datos replicados en tres y hasta ocho sitios, los datos únicos se deben escribir de manera relativamente equitativa en cada sitio. Cuando los datos se escriben en cantidades iguales entre los sitios, cada sitio tendrá una cantidad similar de fragmentos de réplica. La cantidad similar de fragmentos de réplica en cada sitio conduce a una cantidad similar de operaciones XOR que pueden ocurrir en cada sitio. La máxima eficiencia del almacenamiento de múltiples sitios se logra reduciendo el número máximo de fragmentos de réplica almacenados mediante XOR.

Tabla 10 Sobrecarga de protección de almacenamiento

N.º de sitios en RG	EC de 12+4	EC de 10+2
1	1.33	1.2
2	2.67	2.4
3	2.00	1.8
4	1.77	1.6
5	1.67	1.5
6	1.60	1.44
7	1.55	1.40
8 (N.º máx. de sitios en RG)	1.52	1.37

10 Conclusión

Las organizaciones enfrentan cantidades cada vez mayores de costos de almacenamiento y de datos, especialmente en el espacio de la nube pública. La arquitectura de escalamiento horizontal y distribuida geográficamente de ECS ofrece una plataforma de nube en las instalaciones que escala hasta exabytes de datos con un *Costo total de la propiedad* que es considerablemente menor que el almacenamiento de nube pública. ECS es una excelente solución debido a su versatilidad, hiperescalabilidad, funciones eficaces y uso de hardware genérico.

A Soporte técnico y recursos

Dell.com/support se centra en satisfacer las necesidades de los clientes con servicios y soporte comprobados.

Los videos y documentos técnicos de almacenamiento proporcionan pericia que ayuda a garantizar el éxito de los clientes en las plataformas de almacenamiento Dell EMC.